DESIGNING NARRATIVE-BASED AI ASSISTANCE FOR SENSEMAKING IN COLLABORATIVE ENVIRONMENTS: CASE STUDIES IN EDUCATION AND DEMENTIA CARE

A Thesis
Submitted to the Faculty
in partial fulfillment of the requirements for the
degree of

Doctor of Philosophy

in

Engineering Sciences

by Dylan Edward Moore

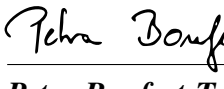
Thayer School of Engineering
Guarini School of Graduate and Advanced Studies
Dartmouth College
Hanover, New Hampshire

March 2026

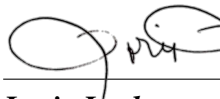
Examining Committee:



(chair) ***Elizabeth L. Murnane***



Petra Bonfert-Taylor



Lorie Loeb



James Landay

Carrie H. Colla, Ph.D.
Dean of the Guarini School of Graduate and Advanced Studies

© 2026 Dylan Edward Moore

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

To view a copy of this license, visit
https://creativecommons.org/licenses/by/4.0/

ABSTRACT

This thesis addresses a gap in the human-computer interaction literature regarding the design, development, and evaluation of narrative-based AI assistance for collaborative, complex problem solving. I explore this design space through three case studies across the domains of education and dementia care. This work encompasses multi-year industry partnerships and longitudinal fieldwork, user-centered design, dataset curation, model training, and system evaluation.

Specifically, the first case study considers a story-based web platform for teaching AI literacy through peer-generated, personalized narrative scaffolding. Learners on the platform showed significant knowledge gains and other learning-related outcomes. To describe the novel design of this system, I introduce the concept of *narrative-based learnersourcing*.

The second and third studies consider introducing AI assistance into under-resourced long-term care facilities to help caregivers provide better-quality care to residents living with dementia. Through a person-centered care approach, I show how an AI system can assist in observing and sharing relevant narrative-based care insights, such as musical interventions, pertinent life history, and uncovered preferences. Following the collection of a custom training dataset, I deployed CareInsights, an intervention that provides caregivers with real-time, chat-based support in the form of context-appropriate, narrative-form descriptions extracted from aggregated video, text, and integrated electronic health records data. I later extended this work, in the third study, to consider a broader range of interface designs and data sharing between families and staff. Across the two care studies, findings indicated outcomes consistent with more effective person-centered dementia care, deepened family involvement in care, and no significant change in measured workload. These studies also surfaced inherent tensions around automation, oversight, and privacy.

In the Discussion, I analyze these case studies' empirical findings through the theoretical lens of sensemaking, which describes the interpretive processes—involving individual cognition and social interaction—through which people organize ambiguous circumstances

into actionable narratives. Bridging theory and practice, I conclude by introducing a novel design framework called *Collaborative Narrative Scaffolding* to guide the design of future AI assistance for complex problem solving in real-world, collaborative settings such as these.

Acknowledgements

This thesis and my graduate school career would not have been possible without support from mentors, colleagues, friends, and family. Specifically, I would like to thank:

- My advisor, Liz, for her guidance, support, and encouragement throughout the course of this research. I knew I wanted to work with you since we first worked together at Stanford and I am so grateful to have had the opportunity to do so. It's been such an incredible experience to watch the Empower Lab grow.

- James, Lorie, and Petra for believing in this work, for being on my thesis committee, and for providing valuable feedback along the way. I know there are many demands on your time, and I really appreciate your support.

- The Memcara team, Christina, Dio, and Brian, for their many, many hours of work on these two field studies and the incredible thoughtfulness that they bring to the important music therapy work they do.

- The PhD Innovation program at Dartmouth for providing funding and support for this research, and Eric Fossum for his guidance and support.

- Songyun, for being an awesome collaborator and friend. I am grateful for your help and support throughout this process.

- My lab mates for their kindness, support, and friendship. I am grateful to have had the opportunity to work with such a talented group of people.

- My friends and family for their love and support.

- My sister, for being my role model and generally just such an amazing, lovely person. You can do anything!

- My parents Carol and Norman for their unconditional care and love. You are both an inspiration to me.

- My wife Melody for always being there for me, including through four years of long distance. I'm extremely grateful to have her in my life.

In keeping with the themes of this thesis, I used AI tools to provide adversarial reviews of my writing and minor stylistic suggestions.

Contents

List of Abbreviations **xv**

1 Introduction **1**

2 Related Work **6**
2.1 Collaborative Work and Systems Thinking 6
2.1.1 Complexity 7
2.2 Narrative and Sensemaking 8
2.2.1 Narrative as a Form of Knowledge 8
2.2.2 Sensemaking 11
2.3 Sensemaking in HCI 14
2.3.1 Storytelling in HCI 14
2.3.2 Scaffolding 16
2.3.3 AI in Deployed Settings 18

3 Teaching AI Through Narrative-Based Learnersourcing **21**
3.1 Introduction and Motivation 21
3.2 Background and Related Work 24
3.2.1 Digital Storytelling for Education 25
3.2.2 Learnersourcing 26
3.2.3 Pedagogical groundings of AI learning 28
3.3 Formative Design 31
3.3.1 Case study 1 (C1): Prototype development and narrative-based learnersourcing scaffolds 31
3.3.2 Case study 2 (C2): Collaboration, cultural narrative, and the birth of Apricot Stone City 34
3.3.3 Case study 3 (C3): Cross-cultural scalability and building on prior contributions 36
3.4 System Design 38
3.4.1 Pedagogically-Grounded Design Strategy 38
3.4.2 Learner Roles and User Journey 40
3.4.3 Story Explorer 42
3.4.4 Story Architect 44
3.4.5 System Facilitator 47
3.4.6 Technical Implementation 50

3.5 Evaluation . . . 52
3.5.1 Participants and Procedures . . . 52
3.5.2 Data Collection . . . 53
3.6 Results . . . 57
3.6.1 AI Knowledge Gains (RQ1) . . . 57
3.6.2 Interest and Engagement (RQ2) . . . 59
3.6.3 User Experience (RQ3) . . . 63
3.6.4 Gender Differences . . . 66
3.6.5 Platform Scalability . . . 70
3.7 Discussion . . . 73
3.7.1 Contributions and Reflections . . . 74
3.7.2 Community Partners and Culturally-Situated Narratives . . . 74
3.7.3 Narrative-Based Learnersourcing for Inclusive Learning Communities . . . 75
3.7.4 Narrative as Empowerment Tool . . . 76
3.7.5 Entering the Age of AI . . . 76
3.7.6 Limitations and Future Work . . . 77
3.8 Conclusion . . . 78

4 CareInsights: AI Infrastructure for Person-Centered Dementia Care 80
4.1 Introduction and Motivation . . . 80
4.2 Background and Related Work . . . 84
4.2.1 Care Ethics and Person-Centered Practice . . . 84
4.2.2 Sensemaking and Documentation in the Context of Caregiving . . . 86
4.2.3 From Documentation to Sensemaking Infrastructure . . . 87
4.2.4 Person-centered Dementia Care . . . 88
4.3 Approach and Formative Work . . . 90
4.3.1 Clinical Partnerships, IRB, and Participant Recruitment . . . 91
4.3.2 Needfinding: Understanding Caregiving as Information Work . . . 92
4.3.3 Baseline Gaps and Literature Guidance . . . 94
4.3.4 Partnering with Caregivers to Translate Field Insights . . . 97
4.3.5 Passive Video Capture via Existing Therapeutic Platforms . . . 100
4.3.6 Label Schema Development Process . . . 101
4.3.7 Early Feasibility Testing of Event Generation . . . 103
4.3.8 Gold-standard Dataset Creation . . . 105
4.4 System Design . . . 107
4.5 Implementation . . . 108
4.5.1 Capture API: Computer-Vision Pipeline for Video Summarization . . . 108
4.5.2 Visualizing Factors of Engagement . . . 110
4.5.3 Context API: Aggregating and Structuring Health Record Data . . . 112
4.5.4 Synthesis API: Context-Aware, Narrative-Based Clinical Guidance 113
4.5.5 Point-of-Care Guidance . . . 115
4.6 Deployment Study and Evaluation . . . 116
4.6.1 Survey Design and Administration . . . 118
4.7 Findings . . . 118

4.7.1 Capture API: Performance Evaluation . . . 118
4.7.2 Context API: Caregiver Evaluation of Generated Video Summaries 121
4.7.3 Synthesis API: Caregiver Evaluations of Generated Responses . . . 123
4.7.4 Point-of-Care Guidance Evaluation Measures . . . 124
4.7.5 Evaluation Synthesis . . . 126
4.8 Discussion . . . 128
4.8.1 Complex Adaptive Systems Perspective . . . 129
4.8.2 Design Implications for Narrative AI Infrastructure . . . 130
4.8.3 Limitations and Future Work . . . 134
4.9 Conclusion . . . 136

5 Family In The Loop: Enabling Family Involvement in Dementia Care 137
5.1 Introduction and Motivation . . . 137
5.2 Background and Related Work . . . 140
5.2.1 Person-Centered Dementia Care in Long-Term Care Facilities . . . 140
5.2.2 Family Involvement in Care . . . 141
5.2.3 Family-Centered Digital Caregiving Tools . . . 143
5.3 Method . . . 144
5.3.1 Design Framework . . . 145
5.3.2 Clinical Collaborators, Care Settings, Participants, and Recruitment 146
5.3.3 Discovery (Addressing RQ1) . . . 147
5.3.4 Definition (Addressing RQ1 and RQ2) . . . 148
5.3.5 Development (Addressing RQ2) . . . 149
5.3.6 Delivery (Addressing RQ3) . . . 151
5.4 Findings . . . 152
5.4.1 RQ1: What are the experiences of families and professional caregivers of care recipients living with dementia in long-term care facilities, particularly with respect to challenges in communication and family involvement? . . . 153
5.4.2 RQ2: How might we design and build an interactive tool that addresses the needs of family members and caregivers and enhances relationship-building to foster a greater sense of engagement for families of PLWD? . . . 169
5.4.3 RQ3: How does the tool impact families' sense of involvement in care as well as other caregivers' practices? . . . 182
5.5 Discussion . . . 191
5.5.1 Envisioning a Data-driven, Person-centered Care Ecosystem . . . 192
5.5.2 Future Work . . . 195
5.5.3 Limitations . . . 197
5.5.4 Navigating Tradeoffs in Scaling Care Systems . . . 197
5.6 Conclusion . . . 198

6 Discussion 200
6.1 Collaborative Narrative Scaffolding . . . 200
6.2 Design Tensions . . . 205

6.2.1 Context Awareness and User Privacy . . . 206
6.2.2 Particularity and Generality . . . 208
6.2.3 Trust and Doubt . . . 210
6.3 Limitations . . . 211
6.4 Operational Realities . . . 212

7 Conclusion **214**
7.1 Summary of Contributions . . . 214
7.2 Future Directions . . . 216
7.3 A Personal Note . . . 217

A Apricot Stone City: Supplementary Materials **218**
A.1 Example Knowledge Assessment Questions . . . 218
A.2 Example Learner-Made Quiz Questions . . . 220
A.3 Pre- and Post-Study Survey Instruments . . . 222
A.3.1 Questions Asked Pre- and Post-Study . . . 222
A.3.2 Questions Asked Only Post-Study . . . 227
A.4 Post-Scene Feedback Questions . . . 229
A.5 Focus Group Interview Guide . . . 229
A.6 Participant Information . . . 231

B CareInsights: Supplementary Materials **234**
B.1 Caregiver-generated Vignettes from Thematic Video Analysis . . . 234
B.1.1 Adam: Verbal Cues of Memory Recall . . . 235
B.1.2 Birdie: Cognitive and Emotional Engagement . . . 236
B.2 Expanded Capture API Technical Description . . . 237
B.2.1 Identifying and Describing Meaningful Moments . . . 241
B.3 Expanded Synthesis API Technical Description . . . 243
B.3.1 Retrieving and Filtering Relevant Context . . . 243
B.3.2 Prompt Construction and GPT-based Synthesis . . . 243
B.3.3 Post-processing and Citation Validation . . . 244
B.4 Example Outputs from the Synthesis API . . . 244
B.5 Survey Items for Person-centered Care . . . 246
B.6 Care Recipient Demographics . . . 247

C Family In The Loop: Supplementary Materials **249**
C.1 Prototype Gallery . . . 249
C.2 Detailed System Overview . . . 250
C.2.1 Onboarding . . . 250
C.2.2 Reels View . . . 250
C.2.3 Explanations View . . . 253
C.2.4 Essentials View . . . 256
C.2.5 Archives View . . . 259

Bibliography **262**

List of Tables

3.1 Baseline AI knowledge, as measured by the pre-study AI knowledge exam. 55
3.2 Percentage change in AI knowledge, measured by performance on pre- to post-study isomorphic AI knowledge exams. (*p < .05, **p < .01, ***p < .001). 58
3.3 Concept mastery, as measured by AI knowledge exams. 58
3.4 Least squares regression to test how learner baseline factors (self-efficacy, readiness to learn, and a detailed self-assessment of AI comprehension) impact knowledge gains, as measured by pre to post exam scores. 60
3.5 Comparing female and male learners on assessed, self-reported, and logged metrics of learning and engagement. 66

4.1 Breakdowns in documentation and sensemaking create reinforcing cycles of poor care quality and caregiver burnout. 98
4.2 Evaluation metrics by sensemaking stage with justification. 117
4.3 Automated engagement detection performance metrics. 119

5.1 Theme 1: Staff-Centered Caregiving Needs 155
5.2 Theme 2: Family Involvement & Information Needs 163
5.3 Staff-Facing Design Decisions (Referencing Existing Practices) 171
5.4 Family-Facing Design Decisions (Referencing Existing Practices) 173
5.5 Final Pre-Deployment Design Adjustments from Late-Stage Prototyping . . 178

6.1 A mapping of the three case studies into the terms of CNS. For each element of the model, this table shows how each case instantiated it through infrastructuring and the evaluation methods used, with pointers to the sections that describe them. (ASC = Apricot Stone City, CI = CareInsights, FITL = Family In The Loop) 204

A.1 Participant demographics for the three preliminary case studies (C1, C2, C3). 232
A.2 Participant demographics for the final system evaluation study. 233

B.1 Details of care recipients at facilities (F) included in all components of the final analysis. 248

List of Figures

2.1 Weick's model of organizing, redrawn from Figure 5.1 in *The Social Psychology of Organizing* [Weick, 1979, p. 132]. Chapter 6 later adapts this model when presenting Collaborative Narrative Scaffolding. 13

3.1 A visualization of important activities in the narrative-based learnersourcing system and how these are associated with different cognitive learning levels. The four activities (know and understand AI, use AI, evaluate and apply AI, AI ethics) are adapted from Ng et al.'s formulation of Bloom's cognitive taxonomy for AI literacy [Ng et al., 2022a]. The figure's visual design is based on the revised edition of Bloom's taxonomy [Anderson and Krathwohl, 2001]; in particular, the stacked pyramid depiction is avoided, which can inappropriately suggest an overly hierarchical and linear progression through Bloom's levels [Clark, 2020]. 39

3.2 A learner-centric perspective of the narrative-based learnersourcing journey. The diagram is read bottom to top to follow the learner's journey through the three roles (explorer, architect, facilitator), reading the learnersourcing tasks associated with each role from left to right. System components are shown in relation to the specific learning activities and roles they support. Sub-tasks that emphasize real-world engagement are highlighted with dashed borders. 41

3.3 Story explorer UIs: Login (top left), progress (top right), graph view (bottom left), adventure view (bottom right). 43

3.4 Story architect UIs: Story creation and editing (top left), code editor (top right), story scene content graph (bottom left), story wiki (bottom right). . . 45

3.5 Facilitator UIs: Scene feedback (top left), system code (top right), guidebook (bottom left), pull requests (bottom right). 48

3.6 Learners who could successfully make the transition from explorer (level 1) to architect (level 2) tended to then successfully progress through the remainder of the learning journey, eventually becoming facilitators (level 4). 61

3.7 UTAUT2 assessment of user experience based on Performance Expectancy (PE), Effort Expectancy (EE), Social Influence (SI), Facilitating Conditions (FC), Hedonic Motivation (HM), Habit (HA), and Behavioral Intention (BI). 63

3.8 Female learners had a greater improvement in mean score from pre to post AI knowledge exam. 66

3.9 Female learners tended to have higher attentiveness scores compared to male learners. 67

3.10 Engagement time on the platform was relatively similar for female and male learners. 68
3.11 The user experience ratings by female and male learners were relatively similar. 69
3.12 Referrals made and accepted by users on the learnersourcing platform. . . . 72

4.1 A timeline of the CareInsights study. The blue arrows pointing to the right indicate that CareInsights remains actively deployed. 90
4.2 Annotated timeline of a caregiver's day showing common information-related tasks, emotional responses, and observed breakdowns across shifts. Illustrations were hand-drawn by the first author based on field observations. Notably, caregivers share short, memorable stories about care recipients as a form of peer-to-peer support; these stories combine background details with direct guidance and are adapted to the needs of each situation. While helpful, this channel is unreliable due to time pressure, poor record-keeping, and staff turnover. **Without a system to capture, contextualize, and synthesize these small but meaningful stories, their insights tend to remain isolated rather than guiding day-to-day decisions.** 93
4.3 Caregivers use a two-device workflow to label sessions. Recorded videos are accessed on the Memcara secure cloud. Caregivers launch the labeling app by scanning a QR code displayed in the web platform, automatically syncing the labeling interface to the corresponding video session. Annotations are made in real time on a secondary device (e.g., tablet or phone), with engagement factors presented as parallel, toggleable tracks aligned to the video timeline. 106
4.4 CareInsights system architecture, aligned with stages of the sensemaking loop. Each API supports a key aspect of caregiving information work: capturing meaningful observations (Capture API), enriching context from fragmented records (Context API), synthesizing insights into actionable narratives (Synthesis API), and delivering real-time, person-centered point-of-care guidance. 107
4.5 Given a video recording of a therapy session and relevant EHR metadata, the pipeline that supports the Capture API extracts visual and textual features using pretrained base models (DeepFace, MediaPipe, Whisper, GPT-4o). These are passed to a stacked ensemble learner (XGBoost), which generates per-second predictions across five engagement factors: attention, emotion, memory recall, reactions to therapy, and behavioral symptoms. Change point detection is applied to these signals to identify potential Meaningful Moments, labeled using rule-based heuristics and summarized using templated GPT-4o prompts. The resulting output includes time-aligned engagement trajectories, classified moment types, and narrative summaries suitable for documentation or review. 110

4.6 Sample output from the Capture API. Five screenshots from the internal interface displaying a single moment in a caregiver–resident interaction, each showing a different engagement dimension selected by the user. For the same video segment, the system overlays predicted engagement levels and presents an automatically generated Meaningful Moment summary. . . 111
4.7 Overview of the Synthesis API pipeline. A natural language query is used to retrieve relevant documents, construct a context-aware prompt, and generate an explanatory, narrative-form response with supporting citations. . . 114
4.8 Caregiver evaluation of generated Meaningful Moment across three dimensions: accuracy, completeness, and utility. 122
4.9 Caregiver evaluation of Synthesis API's generated responses (i.e., responses to user queries) across three dimensions: accuracy, completeness, and utility. 124
4.10 Pre/post study comparison of NASA Task Load Index scores across six workload categories. Lower scores reflect reduced frustration, cognitive burden, and time pressure. None of the differences in the subscales reached statistical significance (all $p > .05$). 125

5.1 Staff and family needs are deeply intertwined—when one side's needs go unmet, it often increases pressure on the other. *All illustrations in this paper are the first author's hand-drawn interpretations of the findings.* 167
5.2 Despite distinct roles, family and staff share overlapping information practices. By organizing these practices—including the aspirational model of "Reels"—around the types of informational needs they address, common ground across roles is highlighted—revealing opportunities to design shared tools that support both perspectives while respecting their unique contexts. 174
5.3 Typical user journeys across FITL's four views—Essentials, Reels, Explanations, and Archives. The flow between views reflects how users navigate escalating informational needs in response to emerging care situations, such as changes in a care recipient's dementia-related behaviors. Feedback can be contributed at any stage and is continuously integrated into the shared data layer, enabling all views to dynamically reflect updated understanding, corrections, and newly observed details. 179
5.4 Excerpts from screenshots of FITL Views, see Appendix C.2 for a detailed overview. 180

6.1 Collaborative Narrative Scaffolding (CNS) is an adaptation of Weick's organizing model, reproduced and discussed more in Figure 2.1 [Weick, 1979, Fig. 5.1, p. 132]. As this chapter explains, this adaptation describes how scaffolding based on a shared narrative record can support enactment. The dashed arrow is the scaffolding path I add. 203

B.1 Excerpt of the caregiver-generated timeline for Adam's session, showing expert annotations of the "Memory Recall" dimension. 236

B.2 Excerpt of the caregiver-generated timeline for Birdie's session, showing expert annotations of "Emotion Pleasant," "Reaction Talking," and "Reaction Singing." . . . 237

C.1 Excerpts from early low-fidelity prototypes. All user content shown here is mocked data. . . . 249
C.2 Excerpt from medium-fidelity prototypes. All user content shown here is mocked data. . . . 250
C.3 A screenshot of Reels View. All personally identifiable information has been pixelated. . . . 251
C.4 Explanations View. All personally identifiable information has been pixelated. . . . 253
C.5 A screenshot of Essentials View. All personally identifiable information has been pixelated. . . . 255
C.6 A screenshot of Archives View. The header and the "Add Care Note" button are not shown (to capture more of the other features in-frame), but appear at the top of the page. All personally identifiable information has been pixelated. . . . 258

List of Abbreviations

Abbreviation	Expansion
ADL	Activity of daily living
AI	Artificial intelligence
AI4ALL	AI for All Open Learning Curriculum
AIM	Acceptability of Intervention Measure
API	Application programming interface
ASC	Apricot Stone City
CNS	Collaborative Narrative Scaffolding
CSCW	Computer-supported cooperative work
EAL	English as an additional language
EHR	Electronic health record
FIM	Feasibility of Intervention Measure
FITL	Family In The Loop
HCI	Human-computer interaction
IAM	Intervention Appropriateness Measure
IRB	Institutional Review Board
LLM	Large language model
MEAM	Music-evoked autobiographical memory
MoCA	Montreal Cognitive Assessment
NASA-TLX	NASA Task Load Index
NPS	Net Promoter Score
PAMI	Person-Attuned Musical Interactions
P-CAT	Person-centered Care Assessment Tool
PLWD	Person living with dementia
RAG	Retrieval-augmented generation
RQ	Research question
TPACK	Technological, Pedagogical, and Content Knowledge
UES-SF	User Engagement Scale, Short Form
UTAUT2	Extended Unified Theory of Acceptance and Use of Technology
UX	User experience
VC-IOE	Video Coding–Incorporating Observed Emotion

Chapter 1

Introduction

Recent advances in artificial intelligence (AI) are rapidly changing computer-supported cooperative work (CSCW) [Shi et al., 2025, Morris et al., 2024], including how people carry out their work [Lee et al., 2025, Noy and Zhang, 2023], how they communicate and collaborate [Hancock et al., 2020, Sharma et al., 2023], and how they find information [Kaiser et al., 2025, Zhou and Li, 2026]. Driving this paradigm shift are foundation models, machine learning models trained on expansive datasets that can be adapted to a wide range of downstream applications [Bommasani et al., 2021]. In particular, a subset of foundation models called large language models (LLMs) have been reported to sustain open-ended dialogue with users, generate functional code, and accomplish multi-step tasks requiring complex reasoning, taking on interactive roles that previously required human cognition [Bubeck et al., 2023]. When these models are incorporated into technical systems that are sufficiently autonomous, capable of understanding their environments, and able to use digital tools to achieve specific goals, the resulting systems are often referred to as AI agents or agentic systems. In this thesis, I examine the design of narrative-based AI assistance for *sensemaking*—the interpretive process through which people organize dispersed and ambiguous circumstances into an account that lets them derive meaning and take action [Weick, 1995, Klein et al., 2007]—in complex, collaborative environments. I approach

this through the challenges and opportunities that emerge as AI agents enter CSCW environments, and how their participation affects that process. The interventions I present in this thesis are on the lower end of what may be considered "agentic" in terms of their independent capabilities without human involvement.

Sensemaking developed in the late 1960s in organizational studies and social psychology and has since been applied, and independently developed, across a number of different fields including Human-Computer Interaction (HCI) [Weick, 1969, Turner et al., 2023, Russell et al., 1993]. In collaborative settings, sensemaking often involves internal and social storytelling processes, as people integrate theoretical knowledge, real-world data, and lived experiences into narratives[1] with practical utility. This utility, as conceptualized in sensemaking theory, serves to reduce equivocal and ambiguous situations [Weick, 1995, Weick et al., 2005], provide a basis for personal and communal identity [Lave and Wenger, 1991, Wenger, 1998], and guide collaborative work [Brown and Duguid, 1991, Orr, 1996].

HCI research has a rich tradition of exploring how technology can augment sensemaking and recent work in this trend has begun examining what LLMs may offer in this regard. However, as I more clearly define and discuss in Chapter 2, there is a gap in examining AI-assistance for complex problem solving at a systems-level. For instance, the National Academies' most recent consensus report on human-AI teaming (published during the formative stages of my case studies) characterized existing research as "severely limited" in its conceptualization of the functions, metrics, and performance-process outcomes involved in dynamically evolving, distributed, and adaptive collaborative tasks [National Academies of Sciences, Engineering, and Medicine, 2022]. The field has advanced since that report appeared, yet a scoping review of human-centered AI research reported that only 9 of 101 papers studied one AI supporting many humans, with a further 13 studying human-AI

[1]Throughout this thesis, I use the terms *narrative* and *story* interchangeably as broad terms for the chronological ways through which people interpret and derive meaning from experiences. Although some scholarly traditions distinguish between these terms, this broader usage has a precedent in narrative research I extend [Turner, 2025, p. 526] and provides a common vocabulary across the case studies while still aligning with Weick's description of a "good story" [Weick, 1995, pp. 60–61], which I discuss more in Chapter 2.

teams [Comes, 2026]. This thesis provides empirical research to help map the design space of AI-assisted sensemaking in dynamic, collaborative settings.

To ground this investigation, I selected two very different domains—education and dementia care—and within them problems where prior literature highlighted salient user needs and clear opportunities for AI-driven narrative support. Specifically, I consider: AI literacy education in a collaborative learning community (Chapter 3, which introduces the deployed system Apricot Stone City) and person-centered dementia care in under-resourced long-term care facilities (Chapters 4 and 5, which introduce CareInsights and Family In The Loop, respectively).

Across these settings, this thesis presents three deployment case studies grounded in longitudinal fieldwork, user-centered design, systems building, and mixed-methods evaluation. Because these case studies remain largely in their published forms, with minor edits and corrections, each retains its domain-specific terminology. However, as detailed below, the theory in Chapter 2 provides the unifying lens for the cross-case synthesis presented in Chapter 6.

Summary of Contributions

- **Apricot Stone City** [Moore et al., 2024], presented in Chapter 3, is a narrative-based learnersourcing platform[2] for AI literacy education. On Apricot Stone City, students learn through interactive, culturally grounded stories and author new educational experiences with AI assistance. The case study found that the system promoted learning outcomes, including knowledge gains, student self-efficacy, and interest and engagement.

- **CareInsights** [Moore et al., 2025a], presented in Chapter 4: I developed CareInsights through a two-year collaboration with seven long-term care facilities. It is

[2]Learnersourcing is a form of crowdsourcing in which "learners collectively contribute novel content for future learners while engaging in a meaningful learning experience themselves" [Kim, 2015].

an instance of what I term *narrative AI infrastructure*: an AI-enabled infrastructure that supports caregivers who provide care to residents living with dementia by passively recording interactions between caregivers and these residents, enriching them with contextual and clinical signals, and generating narrative-based feedback to support day-to-day person-centered care. CareInsights was integrated into existing electronic health record systems and other caregiver tools and used in real practice. The caregivers who wrote the original documentation rated the narrative-form notes that CareInsights generated as meeting or exceeding the quality of their own, and in using these, reported outcomes consistent with more effective person-centered dementia care. Caregivers also reported no significant change in measured workload from the integration of CareInsights into their workflow.

- **Family In The Loop** (FITL) [Moore et al., 2025b], presented in Chapter 5: One finding from the CareInsights study was that family–facility relationships could be improved by better communication, including story sharing. I developed FITL in response to this need. FITL investigated design considerations for supporting data sharing between families and facilities, building on story-based insights from CareInsights. Specifically, FITL is a web-based interface that synthesizes diverse care data (e.g., video, notes, and records) into contextually relevant, emotionally resonant updates tailored for staff and family members through four interlinked views. Findings from a deployment study indicated that FITL improved staff access to critical resident information and improved family engagement in resident care. Findings also brought into focus important, unresolved questions about data privacy and consent that the chapter discusses.

- **Collaborative Narrative Scaffolding**, presented in Chapter 6: This chapter synthesizes these case studies. While the individual chapters frame sensemaking within domain-specific paradigms—such as pedagogical theories for learning, or geronto-

logical frameworks for person-centered dementia care—this discussion steps back to offer a broader generalization. Specifically, I draw upon and extend the less domain-specific literatures of sensemaking and scaffolding discussed in Chapter 2. With this body of work as a reference, I frame my findings in terms of this novel design framework. I discuss how Collaborative Narrative Scaffolding can help inform the design of narrative-based, AI-assisted sensemaking support for complex problem solving in collaborative settings.

Chapter 2

Related Work

This chapter covers the related work that the thesis builds upon, with more domain-specific work reviewed in each case-study chapter. Section 2.1 describes collaborative work through the lens of systems thinking, and Section 2.2 examines narratives in relation to the theory on sensemaking. Finally, Section 2.3 considers how these traditions have influenced HCI and the present state of AI-assisted sensemaking.

2.1 Collaborative Work and Systems Thinking

Collaborative settings can be understood through the lens of systems thinking, a holistic way of making sense of complex problems by examining the connections, feedback loops, interdependence, and boundaries of a "system" [von Bertalanffy, 1968, Arnold and Wade, 2015]. Donella Meadows, an influential scholar in systems thinking, defines a "system" as "an interconnected set of elements that is coherently organized in a way that achieves something" [Meadows, 2008, p. 11]. In one of her illustrative examples, she points out that a football team can be considered as a system because it remains the same team even after all its players have been swapped out for new people [Meadows, 2008, p. 16]. Systems thinking can be a powerful tool for effecting change in systems because this lens can help identify *leverage points*, or places in a system where small, well-placed interventions can

produce much larger overall effects [Meadows, 1999]. What makes those points hard to find, and their effects hard to predict once an intervention is made, is the complexity of the systems themselves.

2.1.1 Complexity

This thesis conceptualizes "complexity" in two interconnected ways. In the first sense, complexity describes the structural characteristics of the challenges confronting actors in collaborative environments, or "complex problems" [Frensch and Funke, 1995, p. 18]. Specifically, when I refer to complex problems or complex problem solving, I am adopting Dörner and Funke's recent framework on the subject:

> Complex problem solving is a collection of self-regulated psychological processes and activities necessary in dynamic environments to achieve ill-defined goals that cannot be reached by routine actions. Creative combinations of knowledge and a broad set of strategies are needed. Solutions are often more bricolage than perfect or optimal. The problem-solving process combines cognitive, emotional, and motivational aspects, particularly in high-stakes situations. Complex problems usually involve knowledge-rich requirements and collaboration among different persons. [Dörner and Funke, 2017]

Under this framework, complex problem solving appears across many fields—such as education [Reiser, 2004, Jacobson et al., 2019] and dementia care [Kolanowski et al., 2017]—where the underlying principles apply even if the exact terminology varies.

In the second sense, the term is a useful reference to the formal field of complexity science. This field emerged from physics, mathematics, and economics to study self-organizing, non-linear systems. Researchers of complexity science have sought to apply algorithmic and computational approaches to understand how macro-scale patterns emerge from localized micro-interactions [Mitchell, 2009]. As I discuss more in Section 4.8.1, this

work has offered a useful framing for the long-term care settings in my case studies, where the behavior of a facility emerges from interactions among caregivers, residents, families, administrators, and the technologies they use, and where an intervention has to adapt to shifting conditions rather than assume stable ones.

In this thesis, I consider the system-level effects that follow from introducing technology into an existing collaborative setting where people engage in complex problem solving [Selbst et al., 2019]. I am interested in the types of systems where designers face what Rittel and Webber described as "wicked problems." That is, problems where "any solution, after being implemented, will generate waves of consequences over an extended—virtually an unbounded—period of time," consequences that "may yield utterly undesirable repercussions which outweigh the intended [or already-achieved] advantages" [Rittel and Webber, 1973, p. 163].

2.2 Narrative and Sensemaking

This section defines and provides a brief historical overview of the relevant literature on narrative and sensemaking.

2.2.1 Narrative as a Form of Knowledge

Storytelling is a fundamental part of the human experience as a mechanism for interpretation and knowledge sharing [Turner, 2025, Bruner, 1991]. Before the development of writing, societies passed down practical guidance, moral codes, and other components of collective identity through oral tradition [Ong, 1982, pp. 34–35, 41, 139–140]. Modern technology has, of course, added complexity to these traditions, but the centrality of storytelling to sensemaking remains, as I discuss more in Section 2.2.2.

However, before going further, it is important to address, formally: What is a *story*? And also, how much trust should people place in stories as a source of knowledge compared

to other, more structured or propositional forms of knowledge? To answer these questions, a review of narrative analysis literature by Turner [Turner, 2025] traced the development of definitions and usages of the terms "story" and "narrative."[1] Through this, he outlined an enduring debate over the reliability of narrative-based knowledge. By his telling, this begins with the writings of Plato and Aristotle. Plato viewed stories as unreliable representations that appeal to passion rather than reason, warning that narrative should be treated with caution wherever careful judgment is required [Plato, 1892]. Aristotle, by contrast, defined a story as a coherent, temporally ordered sequence of events oriented toward a particular end [Aristotle, 1996, Liveley, 2019]. On Turner's reading, Aristotle held that the point of a story lies in this end, or *telos*, which is the meaning the story conveys to an audience that follows its events through, and from which benefits such as aesthetic pleasure or learning can arise [Turner, 2025, p. 525]. Turner describes how debate persisted into modern social science and philosophy, with these fields retaining a long-standing skepticism toward narrative-form knowledge [Turner, 2025, Ewick and Silbey, 1995], a preference for structured over narrative-form representations that technical systems design has shared [Berg and Bowker, 1997, Norris et al., 2022]. This changed with the "narrative turn," which began in literary theory during the 1960s and expanded into history and the social sciences by the early 1980s as researchers (including the pioneers of sensemaking) began using stories to understand human life and reality [Hyvärinen, 2012, Weick, 1995]. Within this tradition, Jerome Bruner offered a contemporary version of Aristotle's conceptualization of narrative-based interpretation that has influenced later HCI work. Bruner distinguished between the "paradigmatic mode," which verifies truth through formal logic and empirical proof, and the "narrative mode," which constructs meaning from human experience through an intentional selection and sequencing of events [Bruner, 1986, pp. 11–13, 19].

Bruner's modes provide a useful reference for thinking about how people distill and share knowledge through narrative as they solve complex problems together. The work

[1] Like Turner does in his analysis, I use these terms interchangeably throughout my thesis [Turner, 2025, p. 526].

of Lave and Wenger, Orr, and others helps situate this process in collaborative settings. People engaged in shared work may form what social learning theorists Jean Lave and Étienne Wenger term a *community of practice* [Lave and Wenger, 1991]. Such a group shares an ongoing concern and develops common ways of pursuing it through regular engagement [Wenger et al., 2002, Pyrko et al., 2017]. Over time, this collaboration produces what Wenger [1998] characterizes as a *shared repertoire* of routines, vocabulary, artifacts, and stories. The stories in this repertoire provide a means through which members can collectively interpret their experiences [Wenger, 1998, Wenger et al., 2002]. Learning within a practice is also bound up with identity formation, since participation involves becoming a distinct actor within a situated social context. As a result, members develop a personal stake in how the shared repertoire represents their needs and experiences, as well as the identity of the community [Wenger, 1998, Wenger-Trayner et al., 2015].

Julian Orr's classic study of photocopier repair technicians likewise found that narrative-based knowledge sharing is central to collaborative work rather than incidental to it [Orr, 1996]. Orr observed that technicians shared stories about difficult repairs that mixed idiosyncratic details of a particular machine and site with more general knowledge about the model [Brown and Duguid, 1991, p. 47]. These were often more useful than manuals because they preserved contextual details and practitioners' situated reasoning, which more formal organizational records tend to lose [Ackerman and Halverson, 2004, Ackerman et al., 2013]. John Seely Brown and Paul Duguid described this fluid quality of stories as a medium as *flexible generality* [Brown and Duguid, 1991, p. 44]. That is, stories as a medium are "both adaptable and particular" because they capture idiosyncratic details while remaining open to use by different people in new contexts. Other work has echoed these findings [Linde, 2001, Bartel and Garud, 2009]. Turner, whose review of narrative research I introduced above, provides an explanation for how stories accomplish this. He identifies four narrative affordances, namely meaning-making, persuading, sharing, and coordinating. He argues that these affordances arise from both narrative's formal structure

and the social norms governing its use within a particular practice [Turner, 2025]. These themes, and in general the work discussed here, align with the literature on sensemaking. For clarity, I defer to the definitions and understanding of narrative-based knowledge as it exists in this tradition for my cross-case analysis in Chapter 6. Within my case studies, I use the more domain-specific framings of storytelling or narrative-based knowledge within education and person-centered caregiving, since these are recognizable to practitioners in those fields working on those specific problems.

2.2.2 Sensemaking

For the interpretive work examined in this thesis, I use the organizational theorist Karl Weick's theory of sensemaking, which he introduced to organizational science in work beginning in the late 1960s as a way of explaining how people deal with equivocal experiences, such as the complex problem solving I have described earlier [Weick, 1969, Maitlis and Christianson, 2014]. This is Weick's organizational tradition rather than the version of sensemaking that HCI research has developed independently of it, which I turn to in §2.3 and again in Chapter 4. Weick et al. described sensemaking as the process through which people structure experience and turn "circumstances into a situation that is comprehended explicitly in words and that serves as a springboard into action" [Weick et al., 2005, p. 409]. Sensemaking, they argued, "is central because it is the primary site where meanings materialize that inform and constrain identity and action" [Weick et al., 2005, p. 409]. Sensemaking stands against the rational decision models that were then dominant in organizational science, which posited that people begin with clear problems and fixed objectives [Simon, 1955, March and Simon, 1958].[2] Sensemaking instead assumes that people facing complex problems first construct an interpretation of their situation that provides an orienting sense of meaning, before deciding how to respond [Comes et al., 2020,

[2]Those models have since been substantially revised. Naturalistic accounts of decision making, developed from studies of experts working under time pressure in the field, describe people as recognizing and framing a situation rather than comparing fixed alternatives [Klein et al., 2007, Klein, 2008].

Comes, 2026]. Through the narratives people produce to accomplish this, which answer the question of what they are doing and why they are doing it, individuals and groups also come to understand their own commitments and identities [Weick, 1995, p. 18]. In this sense at a practical level, and at some risk of reductionism, I consider sensemaking as a process by which stories are produced and used as a means of organizing, preserving, and communicating useful information. Weick claims that the characteristics of these stories that make them useful are in fact narrative qualities, e.g., plausibility, coherence, and memorability, rather than their accuracy alone:

> If accuracy is nice but not necessary in sensemaking, then what is necessary? The answer is, something that preserves plausibility and coherence, something that is reasonable and memorable, something that embodies past experience and expectations, something that resonates with other people, something that can be constructed retrospectively but also can be used prospectively, something that captures both feeling and thought, something that allows for embellishment to fit current oddities, something that is fun to construct. In short, what is necessary in sensemaking is a good story. [Weick, 1995, pp. 60–61]

The settings in this thesis confront people with both uncertainty (missing information) and equivocality (multiple plausible interpretations) [Weick, 1995, pp. 27, 95]. Weick describes the process people use in these cases as four elements: ecological change, enactment, selection, and retention [Weick, 1979, p. 130], as shown in Figure 2.1.[3]

[3]For simplicity, my figures (Figure 2.1 here and Figure 6.1 in Chapter 6) omit his use of positive and negative path markings [Weick, 1979, p. 132]. This omission is consistent with what later researchers have similarly done [Jennings and Greenwood, 2003, Lu, 2017].

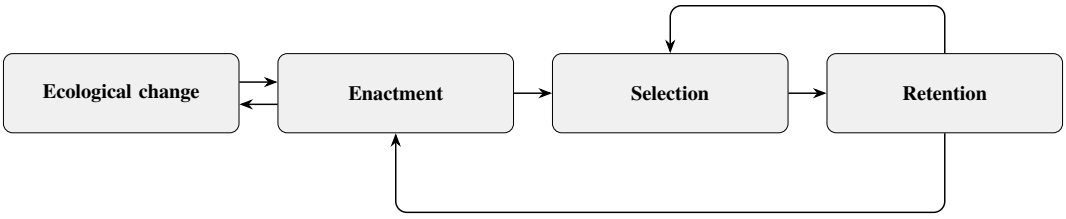


Figure 2.1: Weick's model of organizing, redrawn from Figure 5.1 in *The Social Psychology of Organizing* [Weick, 1979, p. 132]. Chapter 6 later adapts this model when presenting Collaborative Narrative Scaffolding.

The process begins with how a person or group responds to noticeable shifts in their surrounding environment, which Weick describes as "ecological change" [Weick, 1979, p. 130]. Bentley provides a concise definition of the remaining elements in *The SAGE Encyclopedia of Corporate Reputation*:

> Weick describes the basic sensemaking process as enactment, selection, and retention. The enactment stage refers to the actions people take (both what they do and what is done to them). The selection stage involves looking back at an action, choosing which aspects to focus on, and assigning meaning to it based on how one sees oneself and one's environment. The retention stage involves remembering the assigned meaning and using that meaning as a guide for future actions in similar situations. Thus, the sensemaking process is ongoing. When people interpret selected enactments as beneficial, they will tend to retain that interpretation and enact those behaviors again. However, when people interpret enactments negatively, they will tend to avoid or change them in the future.
>
> Sensemaking is a social process. Sometimes, members of an organization work together to make sense of complex situations, as in the case of a new start-up business or a corporate crisis. At other times, sensemaking is performed at an individual level, but it is still affected by other people's words and behaviors.

> For example, a new employee may privately interpret her boss's actions more negatively if her coworkers often complain about the boss.
>
> When people are determining how to interpret a situation, sensemaking guides them to plausible, but not necessarily accurate, interpretations. [Bentley, 2016, p. 743]

Scholars have since expanded Weick's theories into "collaborative sensemaking"—a process requiring participants to negotiate diverse perspectives into a shared understanding [Paul and Reddy, 2010, Convertino et al., 2008]. While useful in other contexts, this framework does not fully align with my case studies. Instead, I utilize Weick's original framing of sensemaking, which inherently accounts for social context and communal infrastructure without demanding a unified consensus among individuals.

2.3 Sensemaking in HCI

In my examination of AI as a core part of my interventions, I draw on Douglas Engelbart's foundational work on how designers might develop technology to augment human intellect [Engelbart, 1962]. Even where HCI research has not called sensemaking out by name, the field has a long tradition of valuing narrative-based practices, and that work overlaps a great deal with the fundamental ideas of sensemaking. Drawing together these perspectives, the case-study chapters' related-work sections review domain-specific examples of such intelligence augmentation that were available to me at the time that each study was published. In the following subsections, I review the more general work that connects sensemaking to HCI.

2.3.1 Storytelling in HCI

In general, HCI research that considers sensemaking has used its own version of this framework because it is helpful for deciding what an interface can do to help someone build a

mental representation that makes a difficult task easier to understand [Pirolli and Russell, 2011, Russell et al., 2026]. Such work began with Russell and colleagues at Xerox, who defined sensemaking as "the process of searching for a representation and encoding data in that representation to answer task-specific questions" [Russell et al., 1993, p. 269]. Across the lineage that followed from this work, the "representations" people develop have often been understood as stories. What counts as a "story" though, can vary substantially, or not be well-defined, depending on the context. Gershon and Page, for instance, define a story by what it does rather than by its form, noting that "a well-told story conveys great quantities of information in relatively few words in a format that is easily assimilated by the listener or viewer," and that people "usually find it easier to understand information integrated into stories than information spelled out in serial lists" [Gershon and Page, 2001, p. 31]. By their telling, a story in this sense need not be a literary text. Relevant prior works in HCI have applied this understanding of a story as being defined by what it accomplishes. For instance, Paul and Morris described "CoSense," a tool for "collaborative sensemaking" that, in one of its views, presented a group's collaborative web search activity as a unified, chronological timeline [Paul and Morris, 2009]. Likewise, Prochaska et al. used the term "deep stories" to describe how stories told collaboratively on social media shaped audiences' interpretations of the 2020 and 2022 U.S. elections [Prochaska et al., 2025].

Recently, HCI work on AI has more explicitly drawn on concepts and terminology from sensemaking to frame LLM-supported information exploration and comparison, as well as AI interpretability [Suh et al., 2023, Liu et al., 2024, Kaur et al., 2022]. The gap that remains, as I described in Chapter 1, is that much of the work closest to the intersection of these ideas has been evaluated in brief, individual-user studies or through bounded chains of artifact reuse, rather than through sustained deployments of narrative-based, AI-enabled infrastructure through which a community's shared record accumulates over time [Suh et al., 2023, Liu et al., 2024, Gu et al., 2024, Fisher et al., 2012]. This pattern holds at the level of the literature as a whole: a review of human-autonomy teaming reports that all the re-

search surveyed was laboratory-based and that most studies used a single human-agent dyad [O'Neill et al., 2022], and evidence for human-AI complementarity still relies heavily on early-stage prototypes or small deployments [Gonzalez et al., 2026]. Related work that comes closer to this model includes Lehmann et al.'s study of AI agents embedded in a collaborative document editor with small teams over one week [Lehmann et al., 2026] and, most directly, Overney et al.'s four-month field deployment of LLM-synthesized composite community stories during a school rezoning process [Overney et al., 2026]. However, the former examines a bounded team-writing task, while the latter presents a fixed set of synthesized stories as a final community-engagement report. My case studies extend this work by treating the shared narrative record as ongoing community infrastructure that accumulates across a real deployment.

I next provide a definition of "scaffolding," a term that originated in the learning sciences and has since been used in HCI work to more precisely describe the generated interface outputs involved in AI-assisted sensemaking [Liu et al., 2024].

2.3.2 Scaffolding

The Soviet psychologist Lev Vygotsky's work on the *zone of proximal development* is a good place to start for defining scaffolding. In the early 1930s, Vygotsky's studies investigated the conditions that were conducive to a student's learning. He described the zone of proximal development as the gap between what a learner can manage alone and what becomes possible with guidance or with more capable peers [Vygotsky, 1978, p. 86]. Scaffolding, in these terms, is temporary support given by another, more capable person to help a student master a task within their zone of proximal development [van de Pol et al., 2010, pp. 271–272]; a meta-analysis of 144 studies of computer-based scaffolding in STEM education found it to be reliably effective across a wide range of designs [Belland et al., 2017]. Wood et al. later offered, without reference to Vygotsky, a commonly used definition of scaffolding as the tutor controlling "those elements of the task that are initially

beyond the learner's capacity, thus permitting him to concentrate upon and complete only those elements that are within his range of competence" [Wood et al., 1976, p. 90]. This description of sensemaking assistance is useful because it identifies the need for the guide to hold both a theory of the task at hand and a theory of the learner, so that support can be personalized and context-specific [Wood et al., 1976, p. 97]. Within the domain of education, it is important for scaffolding to also fade from view, via "the gradual removal of supports until students are on their own" [Collins et al., 1989, Puntambekar and Hübscher, 2005]. Through fading, a learner eventually gains self-sufficiency and advances to more complex challenges instead of remaining dependent. The learning scientist Roy Pea distinguishes scaffolding from distributed intelligence, a term Pea had introduced in earlier work on distributed cognition; distributed intelligence is not meant to fade and rather describes the lasting support of a well-designed environment [Pea, 2004, pp. 431, 441–443] [Tabak, 2004, p. 312].

Within the literature on scaffolding, various authors have discussed the ways in which stories can be used as scaffolding to support learning or problem solving [Pesco and Gagné, 2017, White, 2007, Fabry, 2023, Huang et al., 2026]. Among these, Fabry [2023] and Huang et al. [2026] have independently introduced the term "narrative scaffolding" to describe this type of support. Most recently, Huang et al., through the framework of sensemaking, considered the complex problem solving task facing a single data analyst working with a data visualization tool. They present the tool's interface as a form of "narrative scaffolding," in which text that the analyst writes to describe their thinking about the data becomes an interactive medium for data exploration. As the analyst writes, the system generates semantically aligned visualizations and tracks how the line of inquiry evolves [Huang et al., 2026].

As these works and other related work that I discuss in Sections 3.2, 4.2, and 5.2 indicate, prior works have examined various intersections of AI-assisted sensemaking, narratives, scaffolding, complex problem solving, and collaboration. However, to the best of my

knowledge no prior work has sufficiently described the intersection of all of these. This is precisely the gap that my thesis addresses via the case studies and the introduction of the novel design framework *Collaborative Narrative Scaffolding*.

To do so, it is important to comment on why deployment studies on AI agents have been historically challenging and less prevalent in HCI.

2.3.3 AI in Deployed Settings

The same capabilities that make AI agents versatile can also present challenges and trade-offs for system designers. For instance, LLMs are highly sensitive to changes in their context window.[4] This might not matter much for a sandboxed AI agent with no connection to the open internet, other users, or external tools. In deployed, real-world settings, however, AI agents with a sufficient degree of autonomy can differ drastically from the deterministic, rule-based software that HCI designers have traditionally considered [Amershi et al., 2019]. As deployed agents acquire and use real-world data, call tools, engage with users or other agents, and take autonomous actions, the situations they encounter and create can drift beyond the distribution of inputs and behaviors a designer could reasonably anticipate *a priori*. On the one hand, this ability to reason and react to unanticipated contexts is precisely what makes agents useful, as they can exhibit believable social behavior [Park et al., 2023] and provide meaningful support to people [Sharma et al., 2023] across open-ended situations. On the other hand, agents can go awry in unpredictable ways when given autonomy. Recently, for example, CNBC reported that OpenAI had disclosed models breaking out of a sandboxed testing environment during an internal cyber-capabilities evaluation and exploiting a vulnerability against another company's systems (Hugging Face) [Capoot, 2026]. To prevent such issues, designers can create guardrails that monitor and delimit agent actions.

[4]The tokenized information an LLM conditions on when generating output, comprising base instructions such as system prompts, user prompts, and data gathered through digital tool calls [Mei et al., 2025].

Navigating these tradeoffs is difficult even in the most commonly examined setting of a single user interacting with a single agent, let alone in more collaborative ones. For example, in both cases there is much work remaining to reduce and understand the effects of LLM hallucinations, that is, plausible but incorrect outputs [Ji et al., 2023], which users do not reliably detect and may act on [Bansal et al., 2021, Buçinca et al., 2021].

In understanding how to design AI-assisted sensemaking support for deployment settings, I argue it is useful to turn to a process Pipek and Wulf call *infrastructuring* [Pipek and Wulf, 2009, Lyu et al., 2025]. In infrastructuring, a designer's objectives are to understand and augment the effective aspects of the infrastructure a community already has, work that Pipek and Wulf attribute to users as well as professional designers, and which continues after a system is deployed [Pipek and Wulf, 2009, Star and Ruhleder, 1996, Le Dantec and DiSalvo, 2013]. In health care settings, for instance, this includes fitting new devices and records into established routines [Bossen and Markussen, 2010, Islind et al., 2021]—something I return to in Chapters 4 and 5. Lebovitz and colleagues illustrate why infrastructuring is particularly important for AI-assisted sensemaking interventions. They observed that such tools in clinical AI settings performed well in reported evaluations against expert-produced ground-truth labels. However, in my reading, the tools had issues in local pilot studies because, when developed without an infrastructuring approach, these systems lacked the situated "know-how" required in daily practice [Lebovitz et al., 2021]. Specifically, the ground-truth labels recorded experts' final diagnoses but not the practices behind them, such as reviewing prior images, and all five tools they studied performed poorly when piloted against local radiologists' own judgments [Lebovitz et al., 2021, pp. 1510–1511].

With this in mind, in my case studies, I take an infrastructuring approach that is, at various points in the case designs, described in the terms of co-design and user-centered design (Chapters 3 and 4) or the iterative Double Diamond process (§5.3.1).

In the following three chapters I present those case studies, reproduced with minimal changes in the forms in which they were published,[5] each reviewing the domain-specific work closest to its own setting.

[5]Because they were written separately, the case studies use several compatible phrasings for the ideas expressed here and in Chapter 6, among them "story-based scaffolding" (§3.5.2.4), "narrative-based learner-sourcing" (§3.1) and "narrative AI infrastructure" (§4.8).

Chapter 3

Teaching AI Through Narrative-Based Learnersourcing

This chapter is adapted from work published at CHI 2024 [Moore et al., 2024].

3.1 Introduction and Motivation

As we enter the "age of AI", it is imperative to foster young people's AI literacy by enhancing both their knowledge and skills with respect to emerging technology. By building awareness and understanding of AI, people are empowered to capably and responsibly navigate an AI-infused world, including as part of critically appraising such technology [Markauskaite et al., 2022]. Such competencies also promote one's ability to use and collaborate with AI in both professional and personal contexts [Long and Magerko, 2020]. For example, exposure to computational concepts in high school helps increase later interest in the field, including among students from traditionally underrepresented groups [Google, 2014, Google Inc. and Gallup Inc., 2016]. It also builds students' readiness to leverage AI in future educational and professional careers [Guzdial, 2015] in numerous industries that are expected to have high demands for AI competencies that will exceed the workforce supply [Hupfer, 2020, Randstad, 2023].

Discussion is therefore growing around the need for AI educational resources [Long and Magerko, 2020, Touretzky et al., 2019], and various programs are being launched (e.g., the AI Center for Excellence and MIT's RAISE initiative for Responsible AI for Social Empowerment and Education). However, the overall reach of AI-learning efforts is still rather limited [Zhou et al., 2020]. In practice, it is challenging to integrate AI subject matter into existing K-12 settings for a variety of reasons, including a lack of instructor understanding and comfort level in teaching AI as well as a generally standardized and rigid school curriculum [Su et al., 2023].

Extracurricular programs therefore provide a compelling opportunity to engage broad learner populations, especially in contexts where traditional classroom options may be limited or lacking. For example, my research partner, the TUMO Center for Creative Technologies, is a free extracurricular program for 12–18 year olds that provides interdisciplinary educational opportunities around the globe, such as programs on emerging design, computing, and technology topics. In this chapter, I describe a collaboration with TUMO to investigate: *How might we effectively teach AI concepts to students through extracurricular opportunities that support a variety of learning setups, diverse learner backgrounds, and variable levels of baseline familiarity with AI?* In particular, I was keen to explore strategies that empower distributed extracurricular facilitators to deliver quality learning experiences that otherwise may not be feasible for them to offer more independently.

My co-authors and I specifically pursued an approach based on learnersourcing. "Learnersourcing" was originally coined by Kim [2015] to describe a practice of student learning through the collaborative production of shared learning resources. The term learnersourcing is meant to evoke the related process of crowdsourcing, in which a task is split into smaller tasks that are distributed and completed by a pool of crowd workers [Khosravi et al., 2021]. Learnersourcing does differ from crowdsourcing in a few key ways, namely its motivation and incentive structure. Whereas crowdsourcing conceptualizes participants as "workers" (typically rewarded through paid compensation),

learnersourcing is pedagogically motivated and emphasizes that the learning benefits of generating and engaging with useful content can be intrinsically meaningful [Singh et al., 2022, Weir et al., 2015]. Learnersourcing can lead to formation of a learner community, though this is not a requirement nor always an outcome of a learnersourcing system [Singh et al., 2022].

To scaffold the learnersourcing process, I focused on narrative-based activities. Narratives have shown promise as a learning vehicle by promoting student interest and engagement [Glaser et al., 2009], including in the context of STEM [Ruan et al., 2020] and more specifically computing education, even in very young learners [Dietz et al., 2021, 2023, Ng et al., 2022b]. Using the mechanics of storytelling, abstract concepts can be explained through familiar and intuitive metaphors [Lester et al., 2014], which makes the ideas more accessible, promotes self-regulated learning [Perry, 1998], and overall creates an immersive and engaging experience [McQuiggan et al., 2010, 2008].

The proposed strategy is both significant and novel, specifically in that it addresses a critical need for supporting learner interest and engagement in learnersourcing tasks. The importance of student interest and self-efficacy has been well documented in traditional classroom contexts [Ainley et al., 2002]. Singh et al. [2022]'s review and classification of learnersourcing systems affirms that these psychological ingredients also impact learning outcomes in learnersourcing contexts. In particular, the authors demonstrate that low self-confidence and low interest in tasks are top contributors to a lack of engagement, thereby highlighting that enhancing student motivation is essential for the viability of a learnersourcing system.

As discussed later in this chapter, narrative-based learning activities have been shown to address precisely such challenges related to interest and engagement; however, research has not yet explored their application in learnersourcing given the relatively nascent stage of such systems. In this way, this work responds to calls for additional research on learnersourcing, such as from Khosravi et al. [2023] who point out that "fundamental work" is still

needed for learnersourcing systems to reach their full potential and be ready for large-scale adoption, including to leverage new opportunities for human-AI partnerships in content creation and evaluation.

Using a narrative-based learnersourcing approach, my co-authors and I developed a technology platform that enables a virtuous cycle of content creation, consumption, and peer feedback, with learners iteratively moving up the ranks and taking on increasingly more sophisticated roles that help sustain the educational ecosystem. The formative design phases involved three cohorts of teen learners who helped refine the narrative-based learnersourcing approach and explore its feasibility and efficacy across common extracurricular scenarios. Building on those insights, this chapter reports on the resulting system that instantiates narrative-based learnersourcing as well as a deployment study to evaluate the approach in terms of outcomes related to knowledge gains, attitudinal shifts, and overall user experience.

Specifically, I aimed to examine the following research questions:

- RQ1: Can a narrative-based learnersourcing platform increase knowledge of AI concepts?
- RQ2: Can a narrative-based learnersourcing platform promote interest and engagement with learning AI concepts in an extracurricular environment?
- RQ3: What is the user experience of a narrative-based learnersourcing platform?

3.2 Background and Related Work

This section reviews foundations for this case study, including digital storytelling, learnersourcing, and pedagogical frameworks for AI literacy education.

3.2.1 Digital Storytelling for Education

As the ability to understand and apply AI topics is increasingly seen as an essential skill [Popenici, 2022], inclusive forms of technical education are needed to broadly support AI learning and promote educational equity [DiSalvo and DiSalvo, 2014, Ferguson, 2012]. For instance, creative and playful learning activities can encourage diverse learners' early interest in AI concepts, as highlighted by recent work on designing inclusive educational technologies for AI learning [Druga et al., 2019]. Other research has demonstrated opportunities to support AI learning beyond traditional classroom settings, including in informal learning spaces [Long et al., 2021], via online tools[1], and in family settings [Druga et al., 2023, 2022].

Narrative-based approaches are applicable to all such contexts (e.g., storytelling and other story-based instruction), as stories are a versatile and effective pedagogical tool [Landrum et al., 2019, Szurmak and Thuna, 2013] including when integrated with technology (e.g., digital storytelling) [Psomos and Kordaki, 2012, Quah and Ng, 2022, Sadik, 2008], for students from backgrounds that face educational inequities [Burk, 2000, Litts et al., 2021], and in the context of computing education specifically [Kordaki and Kakavas, 2017, Tengler et al., 2021]. For example, recent research has shown that learners are able to grasp computing concepts and create engaging stories through voice-based interfaces [Dietz et al., 2021] as well as comic-based [Suh et al., 2020] and multimodal (voice and visual) programming environments [Dietz et al., 2023]. Particularly pertinent work indicates that digital story writing can promote AI literacy, including students' ability to move beyond levels of knowing and understanding to actually using and applying AI knowledge to solve real-life problems [Ng et al., 2022b].

Storytelling is recognized as one of the earliest methods of teaching, in large part because it serves as an instrument for distilling complex information into more approachable, familiar, and engaging formats [Pedersen, 1995]. Today, technology further amplifies the

[1] https://www.media.mit.edu/projects/impact-ai-k-12/overview/

immersive nature of stories through interactive audio, video, and images [Robin, 2006]. Digital storytelling has therefore been utilized across educational domains to enhance student motivation, engagement, and ultimate knowledge retention [Dupain and Maguire, 2007, Sarıca and Usluel, 2016, Yuksel-Arslan et al., 2016].

Narrative creation also enables knowledge sharing [Yılmaz, 2017]. By using story-based approaches, learners can leverage their past personal experiences, thereby creating more active participation in the learning process and enhancing learning outcomes [Smeda et al., 2010]. In this way, digital storytelling is often considered an excellent strategy to promote a culturally inclusive environment, by empowering participation from traditionally marginalized voices [Moutafidou and Bratitsis, 2018]. In addition, digital storytelling is often a more intuitive way for students to approach abstract concepts, and it also positively shapes learner attitudes around the educational topic at hand [Bromberg et al., 2013].

By emphasizing creative skills while simultaneously supporting systematic learning processes, digital storytelling further facilitates imagination, idea generation, information gathering and organization, active learning, self-expression, and problem-solving [Dupain and Maguire, 2007, Yang and Wu, 2012]. In these ways, digital storytelling is a natural vehicle for acquiring and cultivating other 21st-century skills in addition to AI learning, namely creativity, critical thinking, open-ended problem-solving, communication, and leadership [Diermyer and Blakesley, 2009, Price et al., 2015, Sylvester and Greenidge, 2009].

3.2.2 Learnersourcing

Learnersourcing is related to crowdsourcing, a technique that engages groups of (typically non-expert) human workers to complete tasks [Kim, 2015]. Crowdsourcing is often done by unpacking tasks into distributable, smaller "micro" tasks that have self-contained context and can be completed by workers in an asynchronous and remote fashion [Estellés-Arolas and González-Ladrón-de Guevara, 2012]. Crowdsourcing has seen adoption across

disciplines [Wazny, 2017], including complex and creative tasks like article writing, decision making, and science journalism [Kittur et al., 2011], as part of the Wikipedia project [Voß, 2005], and in open source software efforts [Olson and Rosacker, 2013].

In 2015, Kim [2015] used the term "learnersourcing" to describe a practice of student learning through collaborative generation of shared learning resources. A key distinction from crowdsourcing is that learnersourcing is focused on promoting learner-centered needs [Weir et al., 2015], rather than leveraging users as a form of labor to meet task requesters' needs. That is, learnersourcing users participate by engaging in activities that not only collectively generate teaching content for other users but also impart learning on a personal level [Singh et al., 2022]. In this model, the crowd of learners both contributes to as well as benefits from created content, resulting in a virtuous learning cycle [Kim, 2015].

Specifically, learnersourcing has been shown to increase learning gains, not only for those individuals who *consume* crowd-created content but also for those *creators* of educational content [Khosravi et al., 2021, Singh et al., 2022]. Recent work has focused on evaluating the quality of such student-generated content [Darvishi et al., 2021] as well as leveraging the method for annotation tasks [Basharat, 2016] and to overcome experts' blind spots when developing content for novices [Guo et al., 2020].

To generate content, learners are prompted to gain an understanding of the subject matter before creating educational artifacts for others, making content creation in a learnersourcing system similar to project-based learning [Almulla, 2020]. Artifact creation involves a process of self-explanation, which has been shown to lead to effective learning [Bisra et al., 2018]. Moreover, content creators actively engage with the learning material as they produce these artifacts, which results in better recall [Crutcher and Healy, 1989] and involves cognitive activities associated with higher-level learning according to Bloom's taxonomy. Additionally, learnersourcing can foster a community where members connect around shared learning goals [Singh et al., 2022]. Further, such community can provide

mentorship and role models while motivating learners to participate and contribute for the good of the collective [Kim, 2015].

Recent work involving learnersourcing emphasizes the growing possibilities for student, educator, and machine partnerships [Khosravi et al., 2023], with ChatGPT emerging as a focal point of student-AI collaboration [Bao et al., 2025, Lo, 2023, Pardos et al., 2023, Singh et al., 2024, Thomas et al., 2023]. Among other considerations, these works explore how ChatGPT can be used to achieve higher quality learner-made content. In the learnersourcing case study presented later in this chapter, I leveraged these insights by taking a human-AI conversational approach to content creation.

3.2.3 Pedagogical groundings of AI learning

With the increasing significance of AI in both professional environments and daily routines, being knowledgeable about AI has come to be framed in terms of "literacy" [Long and Magerko, 2020], as students will need to be literate in AI skills to thrive in a future where such technology is prevalent [Wang and Cheng, 2022]. Research interest in AI literacy education was limited before 2016, but the number of publications on this topic has recently surged [Ng et al., 2021]. This rise in attention, as well as the formation of major collaborations such as UNESCO's Artificial Intelligence and the Futures of Learning Project, mark a recognition that AI technology will increasingly permeate the fabric of everyday life and that the public, including young members of society, require knowledge around fundamental AI concepts.

Various theoretical and pedagogical frameworks have emerged to guide educators in developing and delivering AI literacy curricula. In this research, I build on Ng et al.'s comprehensive review [Ng et al., 2022a]. In particular, Bloom's taxonomy is used for designing learning objectives, activities, and assessments [Bloom et al., 1956]; and Technological, Pedagogical, and Content Knowledge (TPACK) describes how those forms of knowledge enable the successful integration of technology into teaching [Graham, 2011].

Beyond these two AI-specific frameworks, a myriad of other pedagogical approaches further guided this work, both prominent theories (e.g., constructivism [Fosnot, 2013, Applefield et al., 2000], experiential learning [Kolb, 2014, Wurdinger and Carlson, 2010], project-based learning [Kokotsaki et al., 2016, Blumenfeld et al., 1991], problem-based learning [Savery, 2015], applied learning [Ash and Clayton, 2009], community-based learning [Bedri et al., 2017, Farnsworth, 2010], and active learning theory [Bonwell and Eison, 1991, Niemi, 2002]) as well as conceptualizations of AI as a 21st-century skill [Ng et al., 2022a, van Laar et al., 2017, Trilling and Fadel, 2009] and models that are less commonly known yet still pertinent (e.g., rhizomatic learning theory [Brailas, 2020]).

3.2.3.1 Bloom's taxonomy adapted to AI literacy

Bloom's taxonomy, introduced by Benjamin Bloom in 1956, was intended to help educators design a balanced curriculum, learning objectives, and student assessments according to multiple dimensions in the cognitive, psychomotor, and affective domains [Bloom et al., 1956, Adams, 2015]. Though often misinterpreted as overly hierarchical, the taxonomy's "levels" of learning are non-linear and highly integrated [Clark, 2020, Bloom et al., 1956]. Overall, it is generally viewed as a useful framework for educators seeking to craft a dynamic, deep, and adaptive learning experience by interweaving different types of learning and higher level cognitive skills through a variety of tasks [Adams, 2015].

Recently, in *AI literacy in K-16 classrooms* [Ng et al., 2022a], Ng et al. proposed narrowing Bloom's original six levels of cognitive learning (knowledge, comprehension, application, analysis, synthesis, and evaluation) to four AI-specific levels: know and understand AI, use AI, evaluate and apply AI, and AI ethics. *Know and comprehend AI* relates to building basic AI knowledge and comprehension. *Use AI* works to transfer AI concepts from theory to practice. *Evaluate and apply AI* builds critical thinking skills for analyzing and solving problems using AI, which typically requires synthesising prior knowledge. Finally, *AI ethics* evaluates the implications of AI systems in the real world, considering

a diversity of perspectives. As with the original taxonomy, the motivation is to ensure AI curricula devote time for student activities that focus on higher-level thinking, such as AI ethics. Further, Ng et al.'s consolidated taxonomy provides useful specificity and practical value (e.g., concrete guidance for supporting problem-based learning) for AI educators.

3.2.3.2 TPACK adapted to AI literacy

The TPACK framework is another highly influential tool for educators planning curricula [Koehler and Mishra, 2009]. Whereas Bloom's taxonomy focuses on categorizing and structuring educational objectives, activities, and assessments around cognitive processes [Adams, 2015], TPACK focuses on the integration of technological, pedagogical, and content knowledge for effective teaching, emphasizing the complex relationships that exist among these components [Koehler and Mishra, 2009, Blackwell et al., 2016]. Importantly, TPACK incorporates students' learning context, in particular the context-dependent role of technology [Rosenberg and Koehler, 2015].

Ng et al. additionally provide an adaptation of TPACK for AI literacy educators [Ng et al., 2022a] that organizes influential literature (e.g., Touretzky et al.'s "Five Big Ideas in AI" [Touretzky et al., 2019]) along TPACK's three dimensions. The adaptation is designed for practitioners, offering example learning activities and guidance when creating interfaces for learning technology [Ng et al., 2022a]. As described in Section 3.4, learner activities and the user interface were therefore structured based on considerations from these AI-adapted frameworks. This adaptation of TPACK also provides much of the AI literacy curriculum, including educational resources, assessments, and instructional delivery methods. I specifically utilized the AI for All (AI4ALL) Open Learning Curriculum [AI4ALL, 2023], which applies the adapted framework to meet standards from NGSS Engineering[2], ISTE[3], Common Core ELA/Literacy[4], and CSTA[5].

[2] https://www.nextgenscience.org/
[3] https://www.iste.org/standards/
[4] https://www.cde.ca.gov/be/st/ss/documents/finalelaccssstandards.pdf
[5] https://csteachers.org/k12standards/

3.3 Formative Design

The learnersourcing system described in this chapter was not created through a top-down development process. Rather, my co-authors and I engaged in multiple rounds of case study workshops to understand learner needs, build out and iteratively test functionality, and generally understand how the system might be utilized by diverse learners in a variety of extracurricular scenarios. All participants had little to no prior coding experience and were recruited by TUMO, which helped with coordination and provided supporting resources. TUMO also supported communication with families and learners and offered oversight about what data could and could not be collected and how it might be used. To iteratively design, implement, and gather feedback on the system, I undertook three case studies with learners in a variety of extracurricular programming formats offered by TUMO. The first case study focused on exploring learnersourced creation of narratives, the second focused on supporting collaborative engagement, and the third focused on scalability and the potential for learnersourced narratives to be effective and engaging across different cultural contexts. Identifiers for case study participants use the format Cx-y (e.g., C1-1, C1-2, C1-3, C2-1, etc.).

3.3.1 Case study 1 (C1): Prototype development and narrative-based learnersourcing scaffolds

With the first case study, I aimed to build out the initial prototype, develop narrative-based learning content, and explore what sorts of scaffolds were needed to help sufficiently guide learners with little to no prior programming or AI knowledge through a successful learnersourcing experience.

C1 was conducted as an online, 3-week workshop in partnership with TUMO, who assisted in recruiting $N = 18$ learners who ranged in age from 12–17 (9 female, 9 male). All participants had little to no prior coding experience. I served as a live facilitator, develop-

ing lesson plans about fundamental AI concepts including machine learning (supervised, unsupervised, and reinforcement algorithms), classification and prediction, model training, testing, internal representations, neural networks, feature sets, and bias. For each concept, a short lecture, educational handout, and activity were created. In prior years, I had taught multiple other workshops on AI with TUMO and drew on this familiarity and past content. My co-author assisted in facilitation and drew on her professional experiences with these AI topics.

The workshop began with an overview of narrative structure, storytelling techniques, and choose-your-own-adventure style games, where learners explored how a linear narrative could be translated into a graph structure with multiple different possible endings. Using this notion of a "story graph" to characterize a narrative as a set of nodes (story content) and edges (transitions among these story scenes and plot points), we introduced students to concepts of graph structure, first in plain language and then in terms of code.

The remainder of the workshop focused on participants devising their own stories to convey AI concepts. Specifically, I would deliver a lecture on an AI concept and utilize the associated educational handout and activities. Participants were instructed to complete worksheets I had created to develop character profiles and map out plot arcs that weaved in the AI concept. This approach was based on Movement Oriented Design, which offers principles for developing educational multimedia narratives that are emotionally engaging and high quality [Sharda, 2007]. As an example, after the session on k-means clustering, C1-9 wrote a story where a wizard challenges the main character to sort the items in a massive bag. In another story, a tourist photographs a series of statues during a day of sightseeing and writes their names on the back. To demonstrate concepts related to training data and classification, the story then follows the character as she visits a new part of town with different statues of the same historical figures and tries to guess their names based on her photographs.

Next, participants were tasked with implementing their narrative content as interactive functionality using the initial system prototype. Participants, and particularly those with little prior coding experience, benefited from creating narrative-based content in stages (e.g., graph level summary, then wiki, then code), which resonates with literature on digital storytelling in education [Vivitsou et al., 2016]. Participants were enthusiastic about the interactive stories they had made and displayed genuine delight with employing stories as a vehicle for learning. Notably, they actively shared the platform with their friends and family out of enthusiasm to show their work to others.

Regarding design implications for the narrative specifically, I found that participants perceived the stories that blended fiction and personal experiences as the most interesting. Narratives that featured local contexts, such as places in or people from Yerevan, were especially exciting to them. I did observe that participants often struggled to generate story content from scratch, so having prompts that invited them to source stories from everyday experiences was helpful. Further, participants appreciated that the story aspects of the learning experience pushed them in directions where they may have originally lacked self-confidence. That is, a number of learners who rated themselves at the start of the study as strong technically but weak in terms of storytelling appreciated the story aspects by the end of the study, and vice versa for students originally confident in storytelling but feeling less sure of their technical capabilities. Such findings reinforce ideas that narratives are an accessible, approachable vehicle for learning, including about concepts where learners initially feel intimidated.

After C1, the key open questions related to how learners might effectively collaborate and interact with one another on the platform, and whether it is viable for such a platform to enable not only virtual engagement from distributed learners but also in-person engagement from co-located sets of learners. I therefore undertook the next case study to explore these questions in preparation for the final evaluation.

3.3.2 Case study 2 (C2): Collaboration, cultural narrative, and the birth of Apricot Stone City

The second case study focused on supporting collaboration and how emphasizing the integration of personally and culturally significant narrative elements might address learners' difficulties in generating story ideas. C2 took place as an in-person, 3-week workshop with $N = 31$ learners located in Yerevan, Armenia who ranged in age from 15–19 (21 female, 10 male). All participants had little to no prior coding experience. While C1 helped confirm that narrative-based learning content can be exciting for learners to utilize and create, it also exposed a need for additional scaffolding to help these learners better collaborate on these efforts. C2 therefore focused on these interpersonal considerations, along with whether the approach could remain inclusive to diverse extracurricular learning setups, particularly those with in-person components.

Before starting the workshop, my co-authors and I improved support for collaboration on the prototype through better ticket tracking and documentation in the story graph component (mimicking features and processes from popular Agile development tools like Jira). For instance, to manage synchronization between the wiki and codebase, we added a procedure for ownership tagging to the guidebook.

Today, Armenia remains in a precarious national position in the wake of the Second Nagorno-Karabakh War. I observed that this point of national identity was a very important backdrop to a storytelling workshop experience for many students. To create space for these unique and meaningful Armenian stories to enter the project, I encouraged participants to seek "stories of joy" as part of the prompts given for sourcing stories. Resulting stories ranged from memories of peaceful moments during the war, to a treasured birthday celebration with a grandparent, to recollections of a favorite song. As learners adapted these plot points into a larger narrative, I encouraged them to add details that would make the story feel as culturally rich as they desired. Some added their favorite music, some

added photos of the locations they were describing, and some included specific cultural experiences.

Learners then merged their individual stories into a cohesive overarching narrative, which was described initially as a "rhizome" [Brailas, 2020] of narrative possibilities. This organic composition of stories resulted in formation of a new setting, "Apricot Stone City", a fictionalized version of Yerevan, the capital city of Armenia. Learners devised this name and appreciated its multifaceted cultural significance. Specifically, the historic buildings of Yerevan are made of a pink volcanic rock called tuff[6], which gives the city a general coloring similar to ripe apricots. Apricots are indeed a symbolic fruit to Armenians, deeply connected to folklore and a favorite treat. As an example of this cultural significance, Armenia was represented by the song "Apricot Stone" at Eurovision 2010. Further, inside an apricot, there is a pit ("stone") that contains a seed that is edible and a favorite of Armenians. The learners took this as a metaphor for the learning experience, where a sweet and delightful narrative contains hard lessons that eventually give way to a nutritious and fulfilling learning experience.

C2 also examined learner collaboration. I found that participants were able to utilize all of the new collaboration features with negligible confusion, which could be easily resolved by the facilitator or a peer. Further, learners appreciated having structured, systematic processes to follow, and these processes did not inhibit their creativity. The need for content moderation did come up several times during the workshop, such as when participants added age-inappropriate language or other content. I approached this issue by denoting two volunteers who had reached the role of facilitator to be content moderators, charging them with reading through all updates to the wiki at the end of each day and bringing to my attention anything that they deemed problematic. If changes were requested, the moderators would create tickets accordingly; such roles could foreseeably be baked into the platform. In general, I responded to emerging issues like these by building out *processes*

[6] `https://en.wikipedia.org/wiki/Tuff`

on the platform for responsibilities that learners could take up themselves, rather than by directly intervening such as changing content myself.

3.3.3 Case study 3 (C3): Cross-cultural scalability and building on prior contributions

For the third case study, $N = 16$ learners were recruited who ranged in age from 14–17 (8 female, 8 male). All had little to no prior coding experience. C3 focused on whether users could in fact build on each other's contributions to continue expanding out and sustaining the platform, including when learners were from different cultural contexts. The fact that C2 took place in Yerevan, Armenia and that all C1 and C2 participants identified as Armenian shaped those workshops and the resulting content. Students centered their Armenian identity within the created narratives, including cultural history and points of Armenian pride.

Given this embedding of cultural identity into learners' created content, C3 explored scalability to a different learner demographic, from a different geographical region, with potentially minimal cultural common ground. I was particularly motivated to examine this question given expressions of enthusiasm from C1 and C2 participants in not only developing content with cultural significance but also for that content to be broadly shared. For example, C2-8 shared that in her favorite part of one story, the plot described Armenians as being known as "warm and hospitable people" and how that made her feel good knowing that *"people from other countries will read about it."* She emphasized that this *"made the project more exciting."* C2-2 felt strongly that she didn't *"want Armenia to only be associated with Kim Kardashian"* but instead she wanted *"it to be associated with the history, and the fun stuff too."* A few participants expressed that they were motivated by the cross-cultural possibilities of sharing the project with learners from other countries; C2-29 remarked it could be interesting for those learners who *"might not get another opportunity to learn about Armenia."*

Further, while C1 and C2 lasted 3 weeks, I wanted to understand if learning benefits could be observed after a more brief experience on the platform. C3 was therefore run as a 1-week workshop in Berlin, Germany. Through an onboarding survey, I verified there was very little cultural awareness or connection to Armenia for C3 participants, despite the mutual affiliation with TUMO.

This workshop began with a review of the existing prototype. Specifically, after C3 participants engaged with the platform, they wrote what they felt worked well and not well on sticky notes, which were then grouped using affinity diagramming. From resulting clusters, I created a list of aspects to keep as-is, along with suggested additions and other changes, which were then collaboratively ranked by importance. The same exercise had been conducted at the end of C2 (i.e., on the same prototype that C3 participants assessed), and a comparison of the recommendations from the C2 and C3 participants revealed few differences. Both groups focused on improvements related to the desire for style consistency across narratives, more options on the wiki for architecting out story structure, and better ways to integrate teaching lessons into the narrative content. These results were encouraging, as they demonstrated a consistency between the two groups of learners.

Regarding the narrative, I was additionally pleased to see that the local cultural references were well-received by the C3 learners. In interviews, these participants reported that they were excited to learn more about and participate in stories about Armenia, with this content seen as *"more intentional"* and *"less random"*. These findings indicate that drawing stories from real-world, localized, and cultural experiences tended to result in more compelling narrative content, even for learners from outside those regions and cultures.

In C3, learners' baseline exposure to computing was much lower compared to C1 and C2. In this group of participants, most had never taken a computer science course and found the wiki and its associated processes much more intuitive than working directly in the codebase. That said, within two days, participants were able to contribute to the content and began addressing some of the areas of improvement they had identified in the

initial review. Given C3's condensed timeframe, learners focused on augmenting existing narrative content (e.g., fixing plot holes, making stylistic improvements, and developing small spin-off stories), demonstrating their ability to build on other users' contributions and support a cycle of improvement and growth on the platform.

3.4 System Design

Tying together these pedagogical frameworks and learnersourcing principles, this section begins by offering a conceptual model that applies Ng et al.'s formulation of Bloom's cognitive taxonomy to the context of narrative-based learnersourcing. Building on this model, it then details how the platform scaffolds a user's progression through gamified levels of AI learning. Finally, it describes the functional narrative-based learnersourcing system that was co-designed and evaluated with learners from the demographic groups that TUMO typically serves.

3.4.1 Pedagogically-Grounded Design Strategy

Building off Ng et al.'s formulation of Bloom for AI literacy [Ng et al., 2021], the approach further adapts that model to satisfy the design requirements of a narrative-based learnersourcing system. Specifically, Ng et al.'s groundwork is nicely compatible with a learnersourcing system given (a) it divides engagement into discrete and measurable learning activities that can be delivered to users as learnersourcing "tasks" and (b) it provides a widely accepted assessment scheme that can be used to measure learning. To reiterate, the levels in Bloom's and Ng et al.'s frameworks are integrated rather than overly hierarchical [Anderson and Krathwohl, 2001], and the system's design takes the same approach.

Figure 3.1 illustrates the pedagogical model, designed around learners completing learnersourcing tasks that require conceptual and technical AI problem-solving. Each task involves multiple choice assessments and can also support peer evaluation. Completing a

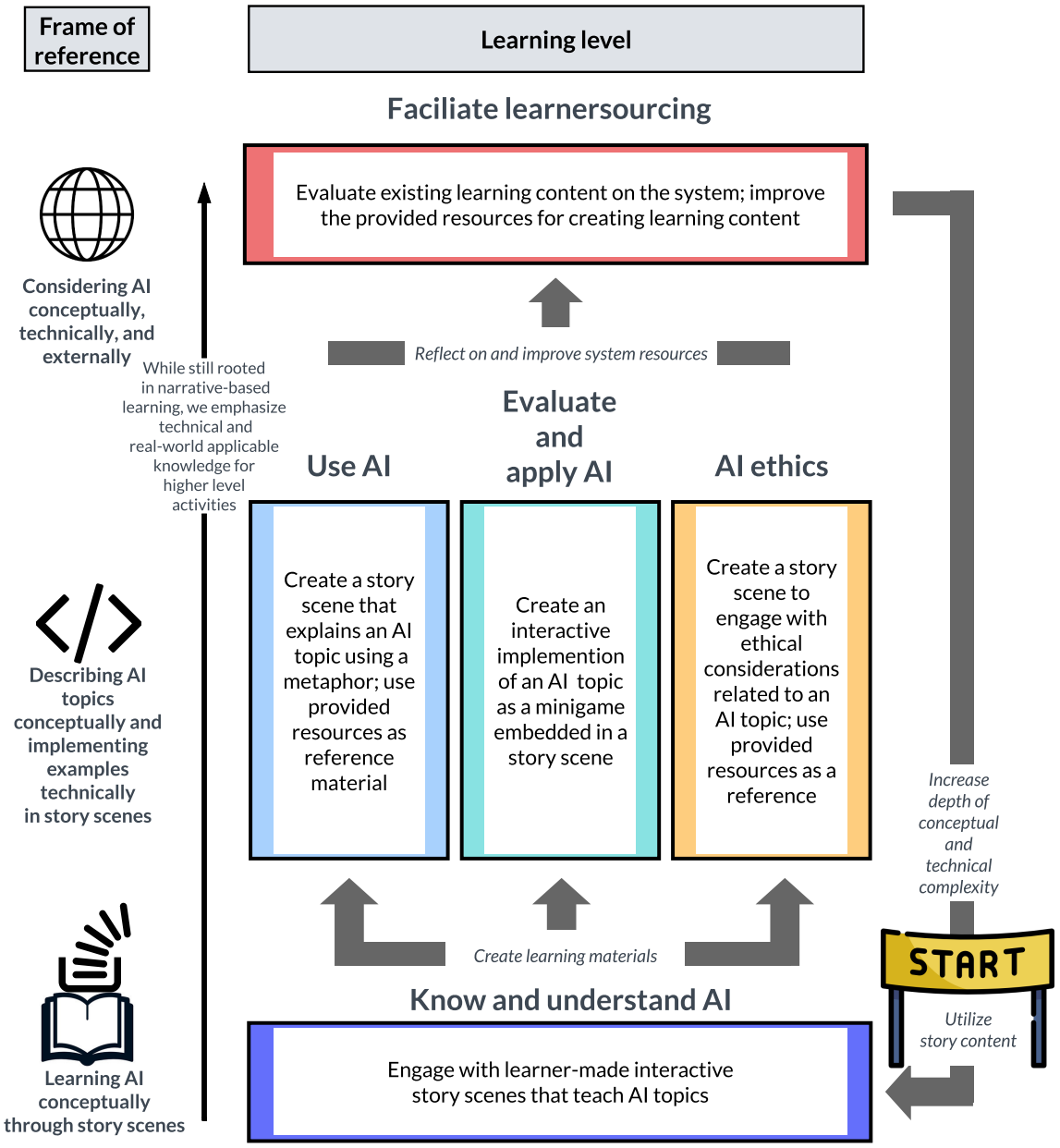


Figure 3.1: A visualization of important activities in the narrative-based learnersourcing system and how these are associated with different cognitive learning levels. The four activities (know and understand AI, use AI, evaluate and apply AI, AI ethics) are adapted from Ng et al.'s formulation of Bloom's cognitive taxonomy for AI literacy [Ng et al., 2022a]. The figure's visual design is based on the revised edition of Bloom's taxonomy [Anderson and Krathwohl, 2001]; in particular, the stacked pyramid depiction is avoided, which can inappropriately suggest an overly hierarchical and linear progression through Bloom's levels [Clark, 2020].

level requires successfully completing a set of tasks that meet specific educational criteria. As learners progress, tasks become more complex and progressively emphasize higher level cognition.

The system is designed for a creative, self-directed learning experience that leverages repeated exposure to AI concepts to promote deeper knowledge comprehension through ongoing story engagement. As learners cycle through levels, they are initially encouraged to simply experiment with activities related to the higher levels, but progressively they come to spend considerable time on more advanced content. The top level ("facilitate learnersourcing") adapts Bloom's original top level ("create") to serve as a learner-enforced check on the real-world relevance of learner-created content across the platform. In this way, the narrative-learnersourcing approach aims to prepare learners for engaging with AI in real life.

Learner levels guide users to reach key curriculum elements by representing that content and associated learnersourcing tasks as components of a game. Upon reaching a higher level, the game objectives change and the learner is rewarded by unlocking access to interfaces that afford increasing rights and roles (e.g., making implementation-level changes to the system), similar to the collaborative model of responsibility found on platforms like Wikipedia.

3.4.2 Learner Roles and User Journey

As learners (interchangeably called users) level up, they become increasingly responsible for building out the system's content and capabilities. Thus, learnersourcing tasks and learner levels are mechanisms for structuring engagement to both deepen a user's learning and support the scalability of the system as a whole. Figure 3.2 overviews the learner experience, mapping Khosravi et al. [2023]'s four core functions in a learnersourcing system (utilize, create, oversight, evaluate) to learner roles. Specifically, each role is defined by a focus on one component of the system: a story adventure component (role = story

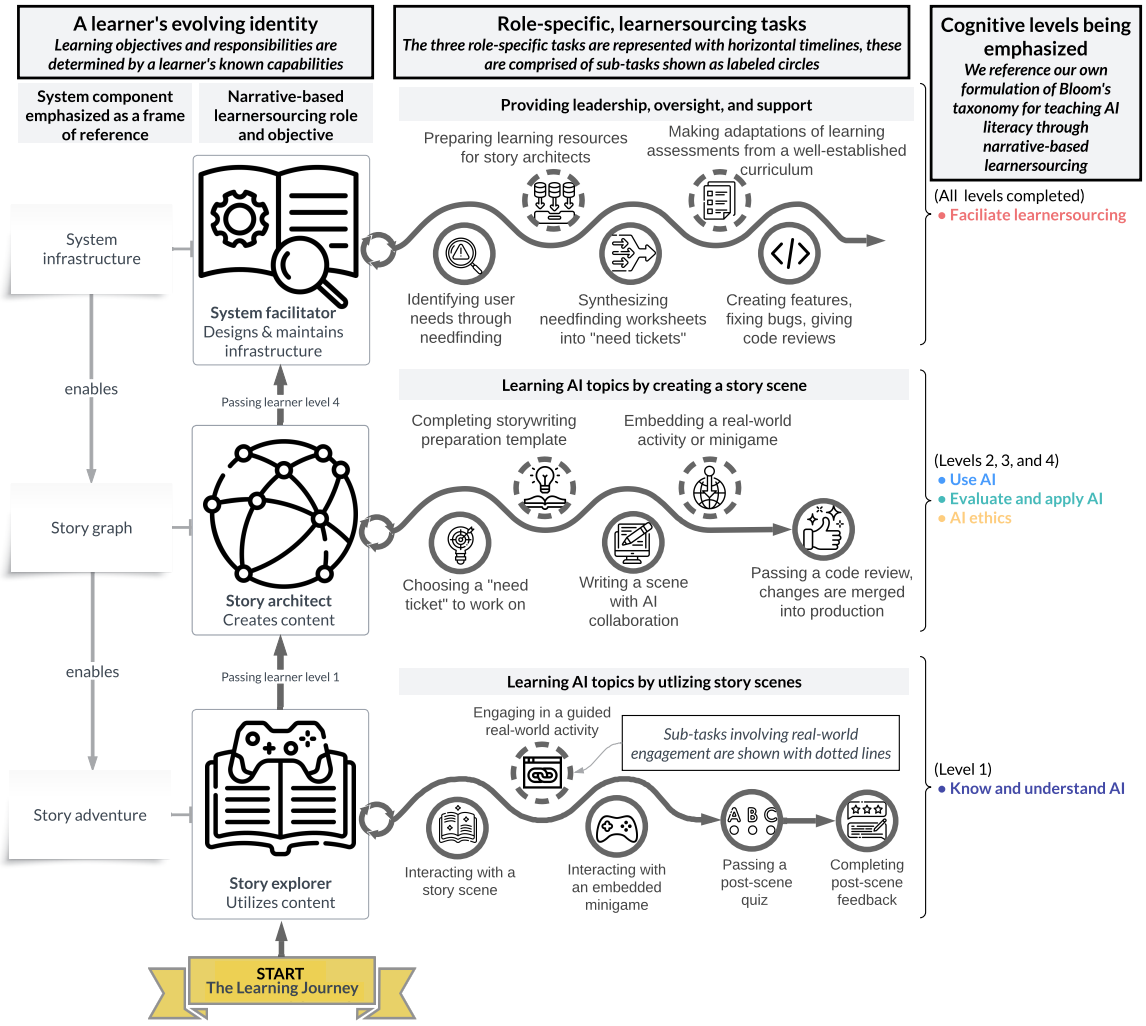


Figure 3.2: A learner-centric perspective of the narrative-based learnersourcing journey. The diagram is read bottom to top to follow the learner's journey through the three roles (explorer, architect, facilitator), reading the learnersourcing tasks associated with each role from left to right. System components are shown in relation to the specific learning activities and roles they support. Sub-tasks that emphasize real-world engagement are highlighted with dashed borders.

explorer, focused on "utilize"), a story graph (role = story architect, focused on "create" and "evaluate"), and story infrastructure (role = system facilitator, focused on "oversight"). Section 3.4.6 describes these components further.

Even though the system provides guided sequences of sub-tasks, an individual's actual learning journey is uniquely personal, steered by narrative preferences, prior technical capabilities, and personal learning objectives. Further, while users are expected to engage in activities they enjoy (e.g., extending a particular plot line in the narrative or implementing mini-game features in code), the system does nudge users to work toward leveling up. For example, after passing level 2, prompts begin to encourage adding a mini-game. The system allows for the involvement of more advanced facilitators at each stage and is designed to promote their engagement and interest so that they continue to provide oversight and contribute to the longevity of the platform.

A learner's progression from explorer, to architect, to facilitator indicates an increasing demonstration of conceptual, technical, and real-world-applicable AI literacy. The following subsections further describe the various roles and associated learning tasks.

3.4.3 Story Explorer

As a story explorer, a learner works to know and understand AI by engaging with existing story content.

"Start": New learner onboarding. A user's first experience on the platform is opening a web URL and landing on the main splash screen (Figure 3.3, top left image). After creating an account, the user enters into the onboarding story scene. This scene explains the interface's controls, and its completion establishes that the user understands and can use the UI. She then transitions to scenes made by other learners, and the story's narratives begin to unfold.

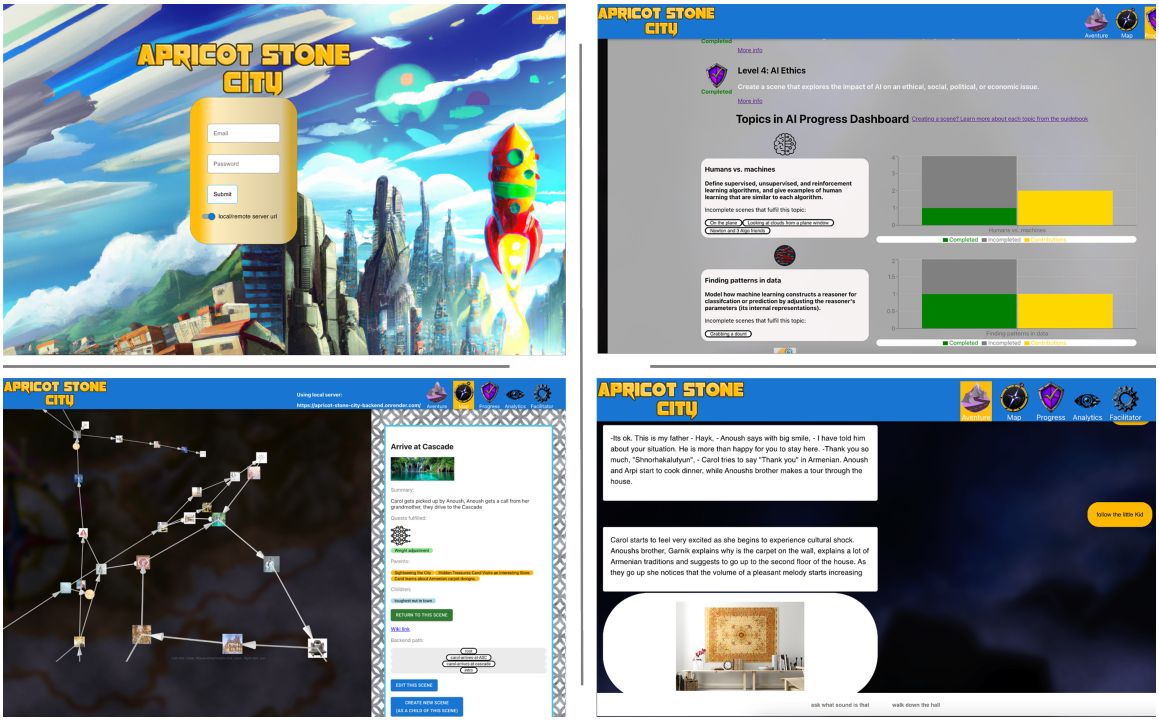


Figure 3.3: Story explorer UIs: Login (top left), progress (top right), graph view (bottom left), adventure view (bottom right).

Interacting with a story scene. The learner receives story content on the web app in the form of text, interactive media, and links. After delivering such content, the app waits for the user to enter free response text or select among hard coded choices to continue the action of the scene. The interface was designed to feel familiar to text messaging UIs like Facebook Messenger or Telegram. Story history is preserved as an infinite scroll. Each scene uses an everyday metaphor to explain a topic in AI (e.g., relating training an AI model to instructing a driver during a taxi ride). These metaphors, as well as the setting and other narrative details that give consistency to the story experience, are generally drawn from a specific real-world place and culture: Yerevan, Armenia. Once the learner has completed a scene, she can revisit it at any time using an interactive graph-like representation of the scenes and their connections.

Engaging in guided real-world experiences. As the action in the story scene unfolds, the narrative links these events to a task that requires the learner to engage in activities

off-platform, either on the internet or in the physical world. For example, one learning task involving training an image classifier involves a user taking photographs in real life.

Interacting with embedded mini-games. Within the context of a scene, the learner can enter into an embedded mini-game. Mini-games support critical thinking and more active engagement with an AI topic. One example on the platform is a variation of "20 questions" that illustrates how the k-nearest neighbors algorithm classifies a data point's grouping. The learners who collaborated to contribute this mini-game (in the story architect role, described in Section 3.4.4) implemented the underlying algorithm using facilitator-provided pseudocode as a reference, and they integrated a narrative backdrop for the mini-game into the flow of the broader story happening in that scene.

Post-scene assessment. At the end of every scene, the learner's comprehension of the AI topic is tested via a multiple choice assessment embedded in the story content. The assessment questions (see Appendix A.2) employ the same extended metaphor as the scene. A learner must correctly answer all questions to progress. Learner responses are logged and used to compute scene-level statistics that are accessible to facilitators as part of system monitoring.

Post-scene feedback. Finally, the learner is presented with a feedback form (see Appendix A.4 for examples) for self-reporting enjoyment and perceived learning during the scene.

3.4.4 Story Architect

As a story architect, a learner focuses on using AI to create new content as well as evaluating and applying AI to synthesize prior knowledge, solve problems, and build critical thinking skills. Figure 3.4 illustrates UIs used by story architects.

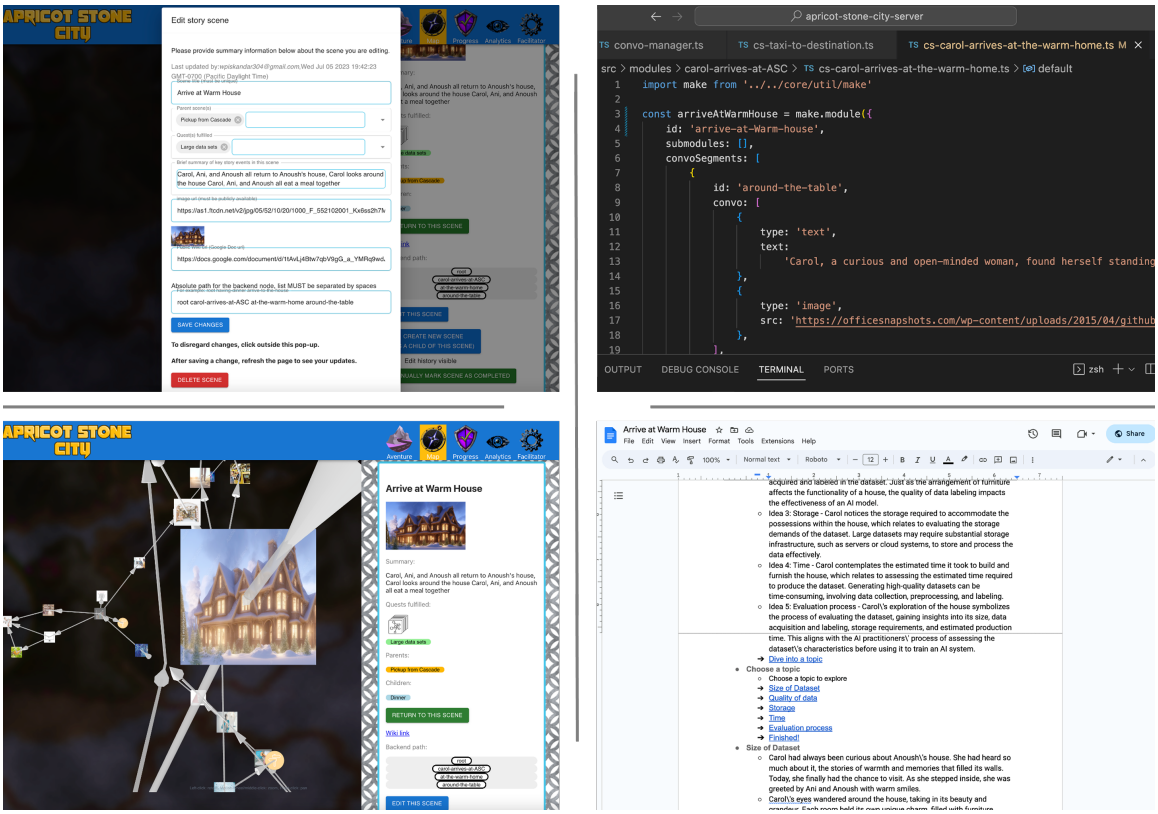


Figure 3.4: Story architect UIs: Story creation and editing (top left), code editor (top right), story scene content graph (bottom left), story wiki (bottom right).

Choosing a "need ticket" to work on. Borrowing from agile development methods, the system uses ticketing to define, organize, and prioritize learnersourced contributions. Each ticket describes a "need" in the story for a scene contribution. For instance, a particular AI topic may require additional explanation if existing content is lacking. After a user makes a selection from a set of available tickets, the system directs the learner through the process of inserting a new story scene or modifying an existing one, as appropriate.

Completing a story preparation template. To prepare for story writing, the system provides a Creator Guidebook (in the form of a Google Doc, which the research team created and that facilitators maintain) of other story scenes and external lecture videos, code examples, and quizzes to help the learner sufficiently master the AI topic her contribution will teach. The guidebook walks the learner through a process of "story-sourcing" to identify real-world narratives that can serve as compelling and relatable teaching metaphors. The guidebook also offers storytelling exercises for architecting an interesting and engaging narrative. To create an outline of her story contribution and embedded quiz questions, the learner follows a story scene template provided by the guidebook. If the learner has questions at any point, she can use a dedicated Discord server to get help from other learners on the platform.

Writing a scene with AI collaboration. The learner next follows a procedure described in the guidebook for converting the completed template into ChatGPT prompts, then reviews, edits, and incorporates the ChatGPT responses into an outline, resulting in a more polished scene script. The learner then integrates the script into the "story wiki," a shared staging environment used to test story transitions and wordsmith. Once the text and media are ready, the learner converts the scene into a TypeScript[7] file using a provided wrapper library and guidebook documentation. She then integrates this file (representing the scene)

[7] `https://www.typescriptlang.org/`

into the server following a linking process outlined in the guidebook. Finally, the learner runs the backend code locally and tests her contribution.

Embedding a real-world activity or mini-game. To complete level 3, the learner must either embed a mini-game or design and incorporate an activity that involves real-world engagement. The guidebook walks through this process. Changes are marked with a placeholder in the wiki and then, after transitioning the wiki to code, added directly as TypeScript code. To complete level 4, the learner must self-educate on an ethical issue related to the AI topic she is working on and create a level 3 type contribution that demonstrates and engages with the issue. The guidebook offers direction by pointing to a variety of information sources, including news outlets (e.g., Forbes and New York Times) and works by prominent scholars in the field (e.g., *System Error* [Reich et al., 2021]).

Passing a code review and merging changes into production. After testing code locally, the learner creates a pull request. Using a built-in messaging service on the platform, she requests that a higher-level learner review the changes. Once reviewed and accepted, the pull request is merged into the production environment.

3.4.5 System Facilitator

As a facilitator, a learner focuses on system oversight and gains increasing responsibility in maintaining and scaling the platform. Figure 3.5 illustrates UIs used by facilitators.

Identifying user experiences and needs on the platform. When acting as a facilitator, the learner monitors the community's engagement with existing content using the analytics dashboard, which includes scene-level statistics such as completion rates, comprehension assessment scores, engagement scores, and anonymized feedback. The learner also monitors and responds to help requests, bug alerts, architect requests, and general complaints on the Discord server. She additionally engages with the story content in a targeted way,

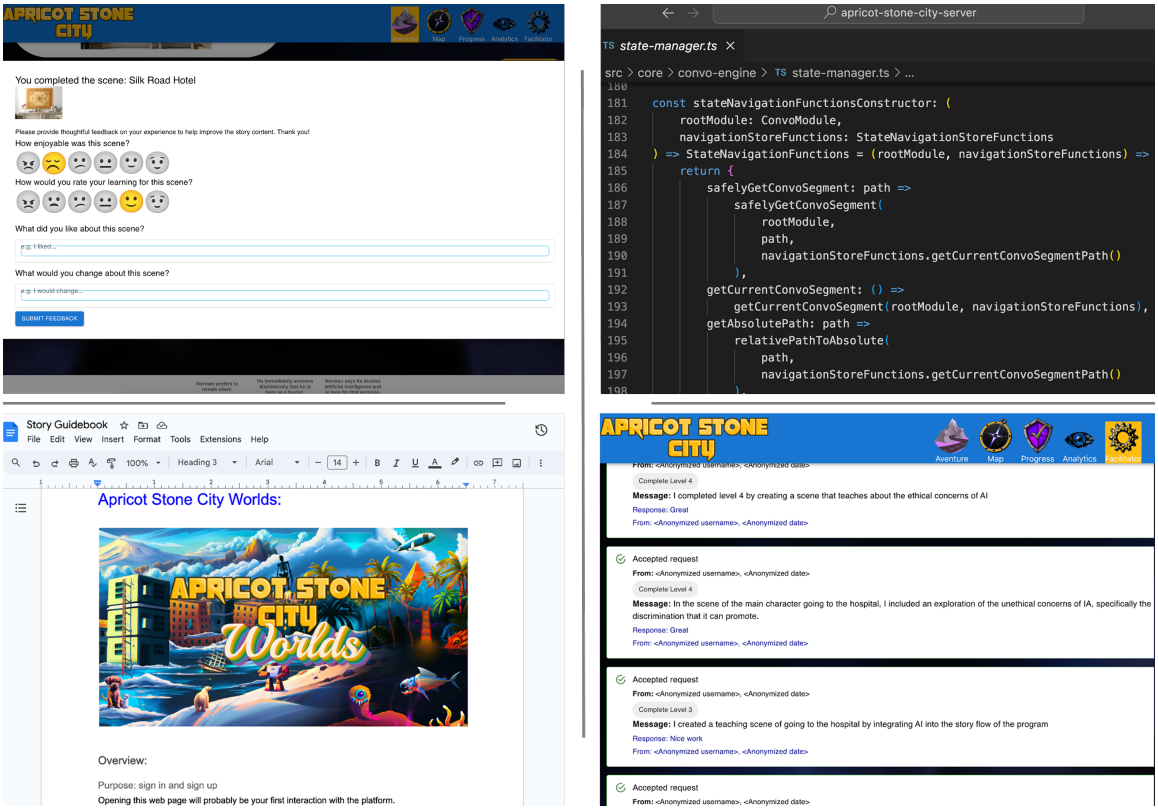


Figure 3.5: Facilitator UIs: Scene feedback (top left), system code (top right), guidebook (bottom left), pull requests (bottom right).

both to hunt for issues (e.g., story gaps in need of content) and to uncover moments of delight [Muldner et al., 2010]. Synthesizing the signals from these different channels, she accumulates her own set of notes describing user needs on the platform.

Preparing learning resources for story architects. Once a facilitator is better aware of issues with story content, she assesses the status of tools and support available to story architects. Specifically, she considers flagged issues in the guidebook, lectures, and platform features, which she addresses by preparing learning resources for architects. The guidebook describes all these procedures.

Synthesizing needfinding results into "need tickets." The learner discusses her needfinding results with other facilitators and helps synthesize findings to compile a set of tickets, which each define a need that should be addressed by an architect. The learner incorporates these into the existing set of open tickets.

Adapting learning assessments from established curriculum. If a need exists for resources for creating post-scene assessments, the learner pulls questions from available trusted online teaching materials (e.g., AI4ALL), modifying them as appropriate.

Creating features, fixing bugs, and performing code reviews. The platform's codebase is available for all learners to view and modify via pull request. When a facilitator identifies a need for a new feature or a bug fix in the underlying system infrastructure (e.g., the web app, server, database, analytics and tracking, or dashboard), she can create and submit a pull request for review by the codebase owners (in this case, the research team, though other administrators are possible, such as extracurricular coordinators or even learners that eventually advance to principal admin status). Facilitators also are responsible for reviewing, commenting on, and accepting or rejecting pull requests.

3.4.6 Technical Implementation

After concluding the preliminary phase of case studies, a final batch of technical improvements was made to the platform to enhance various aspects of the user interface and to instrument the system with logging for the evaluation study. For instance, graph view animations were improved[8], the custom web-app peer-to-peer messaging service was refined, and IdleTimer[9] was integrated for detailed engagement tracking [Doherty and Doherty, 2018] through mouse, keystroke, and React DOM monitoring [Lee et al., 2009]. Multiple rounds of internal bug testing were performed and fixes were implemented accordingly as part of finalizing the system and moving from the "prototype" to "production" phase of the design process.

As described previously, the system consists of three main components: the story adventure, the story graph, and the story infrastructure. The following paragraphs provide technical details about each.

Story adventure component. The story adventure component provides a web application to enable a broad base of learners to engage with story content. Specifically, the component supports state-dependent concurrent sessions, a chatbot-esque choose-your-own-adventure story experience (where choices can be either free response text or multiple choice), progress tracking, and an animated 3-dimensional graph-based story navigation system. These capabilities are enabled by three web views: an "adventure view," "graph view," and "progress view." A discussion of the backend can be found in the story graph component description below.

Adventure view supports an infinite scroll of multimedia, links, and text messages with an interactive display that supports input via multiple choice selections and free text responses. Graph view displays each story scene with a representative picture floating in three-dimensional space (projected using a force-graph calculation). The pictures are con-

[8] Using `https://github.com/vasturiano/react-force-graph`

[9] `https://idletimer.dev/`

nected according to parent-child relationships between the scenes. A learner can navigate the interface with standard, mouse-based zoom, rotate, and translate controls. The learner can select a scene by clicking on its picture to see an interactive pop-up summary. Using this summary, the learner can return to scenes in adventure view (provided they have already completed the scene). The progress view displays information about the learner's learning progress, including progress toward completing level requirements and statistics related to story scene creation and consumption.

Story graph component. The story graph component has two objectives: supporting the story adventure and enabling new content creation. The first is accomplished via an exposed API from the story adventure frontend views. Powering this API is a Node.js server and Firebase database collectively referred to as "the storyteller." The storyteller is a finite-state-automaton that maintains a graph representation of story content and a history of every learner's "story state" as determined by his or her story interactions. It uses these, when provided with a learner's latest interaction, to respond to "scene continuation" API requests. The second objective is a balancing act, as it involves making the barrier to content creation low for novice learners while supporting advanced learners' technical creative expression. This is accomplished by dividing the content creation process into three discrete stages, each involving a progressively lower-level tool; these are the "graph view create GUI," the "story wiki," and the "storyteller content graph."

Story infrastructure component. The story infrastructure component is the underlying codebase and learner-driven design and maintenance processes that enable architects to create content that meets the needs of explorers. This component's primary features are the frontend codebase, storyteller state management and core functionality, analytics view, and guidebook.

3.5 Evaluation

Having built and refined the platform (named Apricot Stone City for reasons described in Section 3.3.2) based on insights from the iterative design process, my co-authors and I next evaluated its ability to promote learning, engagement, and generally positive user experiences.

3.5.1 Participants and Procedures

Participants were recruited through TUMO's existing channels including email lists as well as through referrals from a former TUMO instructor. Snowball sampling was also utilized [Naderifar et al., 2017], encouraging people to share the invite with interested acquaintances.

In designing for a target demographic, my co-authors and I focused primarily on high school aged learners, tailoring the AI curriculum and teaching strategies to this age group [Ng et al., 2022a]. However, from the beginning we were sensitive to concerns from TUMO that, while a target demographic is important, their web-based extracurricular programs include popular offerings that are open access. They found that these are typically utilized by high school aged learners as well as (to a lesser extent) middle school aged learners and college aged students and above. Therefore, to align with the partner organization's vision of broadly accessible extracurricular educational offerings and to support inclusive participation among anyone aiming to gain AI literacy, age-related exclusion criteria were not employed.

The final sample consisted of $N = 27$ participants representative of the learner community that TUMO serves. 14 participants were high school aged (14–18 years old), 11 were college aged (19–22 years old), and the remaining 2 participants were 23+ years old. All participants had at some point been affiliated with TUMO.

In addition to gaining access to the Apricot Stone City platform as described in Section 3.4 and associated resources (e.g., videos, user manual), a Discord chat group was created where learners could interact with each other for peer-based Q&A, collaborative coordination, and general social engagement. The study ran for 1 week, similar to case study 3, as we were interested in exploring the minimum viable timeframe by which positive effects could be observed with respect to knowledge gains, learner attitudes, and sense of community, among other outcome variables described next. The 1-week timeframe is also representative of many of the most popular extracurricular programs offered by TUMO. Procedures were reviewed by the IRB at Dartmouth College.

3.5.2 Data Collection

3.5.2.1 Exams, Quizzes, and Other Learning Assessments

Learning gains were measured through a comprehensive multiple choice AI knowledge exam administered pre-study and post-study. These exams were isomorphic. Within the platform, post-scene AI knowledge assessments also served as a test of understanding for the AI topic covered in that specific scene. The exam and post-scene assessment questions were created based on the learning objectives and "Unpacked" sections of the AI4K12 framework [Touretzky et al., 2019, AI4K12 Initiative, 2020]. Appendix A.1 and Appendix A.2 provide questions from exams and post-scene assessments.

I created the exam, while post-scene questions were created by learners in the architect and facilitator roles, using reference materials that the research team created. During learner assessments and exams, question ordering and multiple choice response ordering were randomized to minimize order effect bias [Wetzel et al., 2016, Lee, 2019]. To discourage random guessing, an "I don't know" option was included [Sherman, 1976] and it was emphasized that scores would be anonymized and that answering honestly would help evaluate and improve the system. Pre-study and post-study exams included thirty questions

(three questions for each of the ten AI topics covered[10]). Post-scene assessments had five questions.

Once learners progress to more advanced levels on the platform, they transition from consuming content to creating it, including quizzes. Those learners' knowledge was therefore also assessed based on expert review of their created content, using a rubric with metrics for the correctness and the helpfulness of content. Specifically, the rubric is based on Denny et al. [2023]'s metrics for comparing the quality of teaching resources generated by an LLM with resources created by students as part of a learnersourcing activity.

3.5.2.2 Automatically Tracked Measures

The platform logs a variety of data, including timestamps of all interactions, visits to platform pages and clicks on specific content, scene edits, and completion of quests and levels. Specifically, affective, behavioral, and cognitive markers of learner engagement [Hu and Li, 2017] are calculated as follows.

- *Sessions*: Session logs track active user engagement in milliseconds and sign in/sign out events. To ensure that only truly active engagement is recorded, IdleTimer is used to track mouse and key events. Session logs also record what component in the UI is being viewed (e.g., "adventure view," "story graph view," etc.)

- *Interaction with content*: All story content (text, images) served to the user and all responses by the user are recorded.

- *Content changes*: All content pushed by users is logged, and users are required to specify the type of change being made as either a "bug fix" or a "feature." Users are also surveyed to self-report the time spent working on content that they push.

[10]The AI topics covered by the platform are: The nature of learning (humans vs. machines), Finding patterns in data, Training a model, Constructing vs. using a reasoner, Adjusting internal representations, Learning from experience, Structure of a neural network, Feature sets, Large datasets, and Understanding bias in datasets [AI4K12 Initiative, 2020].

- *Social interaction*: All communication on the platform between users is logged, using the on-platform messaging capabilities. Requests to the facilitators to fix bugs or review content as well as facilitator replies are also logged.

- *Attentiveness to content*: Attentiveness is operationalized as the amount of time spent on a given story scene, normalized (divided) by the amount of words in that scene.

- *Scenes and topics completed*: A timestamped record of all the scenes and AI topics that a user has completed.

- *Progress through levels*: A timestamped event record of each time a user advances to a new level.

Table 3.1: Baseline AI knowledge, as measured by the pre-study AI knowledge exam.

Mean	**Median**	**Std Dev**	**Min**	**Max**	**S-W stat.**	**S-W p**
38%	37%	23%	0%	97%	0.96	.35

3.5.2.3 Surveys and Self-Reported Data

At the beginning of the study, before administering the knowledge assessment, participants were given a questionnaire to gather personal attitudes about AI-related self-efficacy and sense of belongingness in computing [Leibowitz et al., 2020, Pintrich et al., 1991, Walton and Cohen, 2007], readiness to learn [Hariyanto et al., 2022, Hung et al., 2010, Williams, 2001], as well as subjective enjoyment regarding reading stories and writing stories. A task-value belief scale was also administered, given that the perceived utility of a topic contributes to a learner's interest in it [Conley, 2012, Lishinski and Rosenberg, 2021, Pintrich et al., 1991].

A post-survey at the end of the study asked these same questions and also assessed user experience and perceived usability of the platform via the User Engagement Scale (UES-SF) [O'Brien et al., 2018] and the extended Unified Theory of Acceptance and Use

of Technology (UTAUT2) [Venkatesh et al., 2012]. Participants were also asked to rate the perceived usefulness of specific levels and content in supporting their learning, asked for open-ended reactions to the story content, and asked about sense of belongingness in the Apricot Stone City learning community. Self-assessments delivered at the end of scenes and quests gathered additional self-reported data about self-perceived learning and emotional enjoyment of scenes as well as feedback about what users liked or would change about the scene. Appendix A.3 and Appendix A.4 provide specific questions from these instruments.

3.5.2.4 Focus Group Interviews

At the conclusion of the study, two focus groups were conducted (see the semi-structured interview guide in Appendix A.5). Given that the study was interested in understanding the sense of community that participants perceived on the platform, focus group style interviews were chosen rather than individual interviews. Seven participants (3 female, 4 male) volunteered to take part in the interviews.

Each of the research questions was explored by inviting participants to describe their perspectives on the favorite things they learned and what aspects of the platform experience facilitated or hampered their learning, how and why their interest and engagement in AI improved or diminished during the study, their likelihood of taking AI or computer science classes in the future, their sense of community on the platform, their reactions around the story-based scaffolding, and factors surrounding their intentions to continue or discontinue using the platform after the conclusion of the study.

Interviews were transcribed and qualitatively analyzed together with open-ended self-report data. Specifically, two authors used inductive coding to surface themes and insights, which are used to contextualize the quantitative results reported in the next section.

3.6 Results

This section presents results from the evaluation of the Apricot Stone City system. Main findings are framed around the research questions. Other notable insights are also described, including gender differences as well as results that speak to the platform's scalability.

3.6.1 AI Knowledge Gains (RQ1)

Foremost, I am interested in examining whether a narrative-based learnersourcing approach can promote learning and increase knowledge of AI concepts.

3.6.1.1 Baseline knowledge levels

First, I examine learners' baseline knowledge of the AI topics covered on the platform (see Table 3.1). A Shapiro-Wilk (S-W) test [Robertson and Kaptein, 2016] suggests the data is drawn from a normal distribution. No significant differences are seen in pre-assessment scores related to gender.

3.6.1.2 Changes in AI knowledge

Between the pre-study and post-study AI knowledge assessments, a significant mean increase of 24.2% was observed (see Table 3.2). Regarding outliers, three learners scored exceptionally low on the pre-exam (<3.33%), and two learners scored exceptionally well (>80%). Highlighting and excluding outliers is useful to understand if aggregate results are skewed by the outcomes of learners who scored near zero or near perfect on the pre-assessment. For consistency, any "excluding outliers" notes (in Table 3.2, as well as later results) refer to this same set of outlier participants.

Running analyses without these outliers shows similar results (a significant mean increase of 23.18%). High scoring outliers did improve on average too (mean 91.67% on the

Table 3.2: Percentage change in AI knowledge, measured by performance on pre- to post-study isomorphic AI knowledge exams. (*p < .05, **p < .01, ***p < .001).

Group	Mean change	Median change	Cohen's d
All learners	+24.20% ***	+23.33% ***	1.08
Excl. outliers	+23.18% ***	+21.67% ***	1.28
Female excl. outliers	+29.67% ***	+28.33% ***	2.21
Male excl. outliers	+17.78%	+15.00%	0.86

Table 3.3: Concept mastery, as measured by AI knowledge exams.

Group	Mean	Median	SD	Min	Max	S-W stat.	S-W p
All pre exam	3.30	3	2.8	0	10	0.92	.03
All post exam	6.33	7	2.87	1	10	0.92	.03
Pre excl. outliers	3.18	3	2.11	0	8	0.96	.45
Post excl. outliers	6.27	6.5	2.76	2	10	0.91	.047

post-assessment, compared to a mean of 90% on the pre-assessment). These results suggest that learners with a wide range of prior experience levels could benefit from participating in the learnersourcing platform. Further, statistically significant differences were not observed when splitting participants by race, nationality, or age, suggesting the platform may support AI learning across diverse users.

To examine the assumption that content creation is reflective of knowledge gains, I next separated learners into two groups: those who created content and those who only consumed content. A t-test comparison does show a significant difference (p = .03, Cohen's d = 1.07), with the content creation group having a larger positive increase in knowledge (mean: 31%, median: 30%, std dev: 19%). Regarding attitudinal factors that may impact knowledge gains, a significant correlation between pre-study self-efficacy and pre to post knowledge assessment scores was observed (r = 0.56, p = .04).

In addition, I was curious to understand if learners could reliably self-assess their own learning. Comparing objectively tested and self-reported assessments of AI knowledge post-study, a significant negative correlation was found (r = -0.43, p = .045). This result suggests a potential Dunning-Kruger effect, whereby poor performers overestimate their knowledge and high performers underestimate their knowledge [Mahmood, 2016]. This

phenomenon has similarly been observed in work on narrative-centered, game-based learning environments [Ocay, 2019]. Such findings indicate the importance of objective, expert-made assessments to reliably gauge learning outcomes versus relying solely on learner self-reports.

3.6.1.3 Learning specific AI concepts

To understand how well learners mastered specific concepts, knowledge assessment questions were next grouped by topic. Considering a learner to have sufficient mastery of a topic if she could correctly answer at least two of its three assessment questions, a significant increase in concept mastery from the pre to the post assessments was observed (Cohen's d = 1.09, $p < .001$)—specifically, a mean increase of 3.04 (median: 3.0, std dev: 2.72), with pre- and post-study concept mastery results shown in Table 3.3.

The connection between knowledge gains and progress through Bloom's levels was also considered by conducting an ordered logistic regression between the difference in pre to post exams and the max level completed by a learner. A statistically significant relationship was found with a coefficient of 3.05 and $p = .04$. This relationship is to be expected, so confirming it adds to confidence in the correctness of these two independent measures of learning.

3.6.2 Interest and Engagement (RQ2)

While promoting knowledge gains is a central goal of this work, it is also important to understand learner interest and engagement during the experience, as both are core factors in sustaining participation in learning activities and the mastery of increasingly complex topics over time [Hidi and Renninger, 2006].

3.6.2.1 Interest in AI concepts

Learner interest is a four-phase construct spanning triggered situational interest, maintained situational interest, emerging individual interest, and well-developed individual interest [Hidi and Renninger, 2006].

Early and ongoing interest is signaled by time spent on learning activities. On the platform, learners spent a cumulative 674.2 hours engaging with content. Of all available AI subtopics, learners spent the most time on the most advanced topics—specifically, neural networks (e.g., "Adjusting internal representations", "Weight adjustment," and "Structure of a neural network")—whereas introductory AI concepts (e.g., "Finding patterns in data") saw the least amount of time. In post-study focus group interviews, learners said that they preferred to engage with the topics they were least knowledgeable about out of interest and curiosity.

Further, in moving from content consumers to content creators, learners generally re-engaged with scenes they engaged with and rated favorably as consumers, extending or creating alternative narrative paths for scenes and quests that had originally captured their interest. All learners cumulatively spent the most time creating scenes for quests related to the advanced AI concept "Adjusting internal representations" (115.48 hours).

3.6.2.2 Engagement with the learning process

In the context of education, engagement is commonly characterized in terms of three dimensions: affective, behavioral, and cognitive [Fredricks et al., 2004].

Table 3.4: Least squares regression to test how learner baseline factors (self-efficacy, readiness to learn, and a detailed self-assessment of AI comprehension) impact knowledge gains, as measured by pre to post exam scores.

Baseline factor	**Coefficient**	**p-value**
Self-efficacy	1.66	<.001
Readiness to learn	-1.14	.046
Pre-study self-assessment	-0.12	<.001

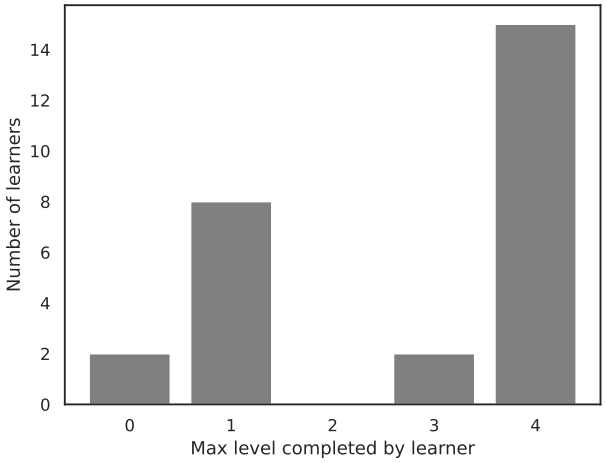


Figure 3.6: Learners who could successfully make the transition from explorer (level 1) to architect (level 2) tended to then successfully progress through the remainder of the learning journey, eventually becoming facilitators (level 4).

Affective aspects of engagement relate to a learner's emotional reactions to the learning activities. The scene that received the highest emotional enjoyment rating (mean: 4.95/6, std dev: 0.92) is called *"Carol, Ani, and Anoush have dinner together."* In this scene, a "foreign visitor" to Apricot Stone City (whose name is Carol) joins a local woman (Ani) and her elderly mother (Anoush) for dinner. Learners' feedback consisted largely of positive comments about the cultural content of the scene—for instance: *"Story content was really pleasing as it showcased some typical traditional Armenian cultural aspects"* (P7) and *"That's a nice change of scene and a great relief after a lot of learning"* (P26).

The main behavioral indicator of engagement is attentiveness to content. Given the attentiveness findings demonstrate gender differences, they are reported in Section 3.6.4 with other such results.

Finally, cognitive engagement is operationalized by completion of learning activities. Specifically, Figure 3.6 illustrates the distribution of max levels reached by participants during the study period. It is important to note that moving from level 1 (story explorer) to level 2 (story architect) entails transitioning from content consumption to creation. The observed drop-off suggests that learners found this challenging. By comparison, the bump

at level 4 (system facilitator) suggests that most learners who were able to successfully make the leap from consumers to creators were then motivated and capable enough to "go all the way." Further, many of these learners had little to no prior coding experience, suggesting the system is able to successfully transition novice learners into roles involving progressively greater concept mastery and platform responsibility.

For most participants, this study was their first time using software development tools like GitHub, VSCode, and React. Based on analysis of interview data, feedback from the post-study survey, and observations of peer-based troubleshooting via the Discord server, getting up to speed with these new tools was a challenge for many learners and a source of pride for those who succeeded. Without such support, learners were at risk of stalling in their progress or even leaving the platform. Such findings highlight future opportunities for needfinding to understand other reasons for drop-off and strategies to minimize it.

3.6.2.3 Personal attitudes and beliefs of learners

Self-efficacy relates to a person's belief in her ability to master a task [Leibowitz et al., 2020, Pintrich et al., 1991]; and readiness to learn refers to the behavioral, cognitive, and socio-emotional skills that indicate preparedness to receive instruction [Millians, 2011]. Given both are important factors in learning, I analyzed how these personal attitudes may relate to the amount of interest and engagement participants demonstrated on the platform, as well as whether the learning experiences afforded by the platform can help to positively shape these attitudes.

A least squares regression was conducted to see how baseline levels of these variables affected knowledge gains, as summarized in Table 3.4. The significant findings suggest that learners with low readiness to learn and low pre-study self-assessment scores (who might therefore typically struggle in both traditional and online learning contexts [Bernacki et al., 2015, Cigdem and Ozturk, 2016]) actually tended to learn more on the platform. I suspect that wrapping the AI concepts into approachable, familiar stories makes the learning pro-

cess more accessible for such learners. Including an additional factor, personal interest in stories, results in a regression coefficient of 5.70, though this relationship is not statistically significant (p = .26).

In addition, evidence was found that the learning experience on the platform can improve personal attitudes and beliefs over time. Specifically, analyzing changes from pre to post responses on attitudinal measures, self-efficacy increased from a median of 10/30 to 21/30 (a 110% change). Readiness to learn remained stable at a median of 16/30, although this is likely due to the nature of the questions in the online learning readiness instrument [Williams, 2001] employed, which may have been less salient or shiftable during this study experience.

3.6.3 User Experience (RQ3)

To understand the user experience (UX) of the platform, UX questionnaires, self-reported sense of community, and how participants reacted to narrative-based learning content were analyzed.

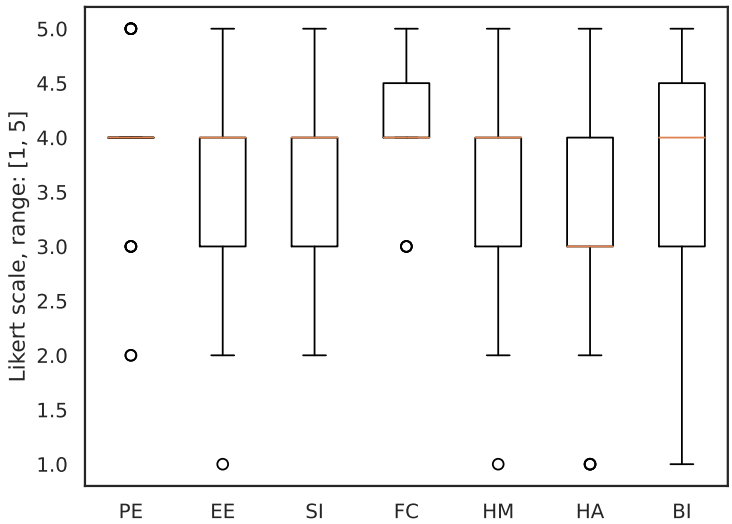


Figure 3.7: UTAUT2 assessment of user experience based on Performance Expectancy (PE), Effort Expectancy (EE), Social Influence (SI), Facilitating Conditions (FC), Hedonic Motivation (HM), Habit (HA), and Behavioral Intention (BI).

3.6.3.1 Semi-quantitative UX measures

To measure how much participants felt absorbed when using the system and how rewarding they found those interactions to be, questions adapted from the UES-SF [O'Brien et al., 2018] were used, on a total scale of 3–15. The analysis indicated a positive user experience on the whole (mean: 11.2, median: 11.0, std dev: 2.5, min: 6, max: 15).

The UTAUT2 [Venkatesh et al., 2012] is another well-established instrument that helps evaluate additional UX dimensions. Figure 3.7 illustrates participants' generally positive responses. The social influence construct demonstrated the highest ratings, suggesting the platform's ability to promote peer awareness and connection even within the study's relatively short timeframe.

3.6.3.2 Sense of community

Further examining social aspects of user experience, participants were additionally surveyed after the study to measure their sense of the community. On a scale of 3–9, results again point to a reasonably strong sense of community (mean: 6.6, median: 7.0, std dev: 1.8, min: 3, max: 9), though not overwhelmingly so. This can likely be attributed to the relatively short study timeframe, and these community ties would be expected to deepen over longer periods of engagement on the platform.

3.6.3.3 Experience with the narratives

As mentioned in Section 3.6.2.2, reactions to the story content seemed to play a main role in participants' emotional engagement with the system. Inductive coding of qualitative data demonstrated that participants consistently felt most confident with AI concepts that were conveyed via analogies built into the story (as reported by P1, P2, P4, P5, P9, P11, P12, P14, P15, P19, P22, and P26). Learners applauded the use of *"metaphor," "analogy," "story examples,"* and *"the process of learning new things by the story."* Participants rarely discussed AI concepts in their feedback outside of references to the storyline and story ex-

planations. Questions around the use of narrative as a tool for AI education overwhelmingly received positive responses.

Although not mentioned as consistently, several participants also expressed excitement for the inclusion of Armenian history and culture in the story. This reflected a trend where scenes with Armenian characters, settings, and subject matter perceived as culturally-authentic also received high enjoyment and learning scores. That said, sometimes Armenian references that wrapped AI content in cultural associations did not always land well with learners. For example, in one scene, a character visits Tsitsernakaberd, the memorial above Yerevan dedicated to the 1.5 million Armenians who perished in the Armenian Genocide. Here, one participant criticized that the location was only superficially discussed, and similarly noted that comparing AI machines to characters with culturally-significant roles could sometimes feel dehumanizing—*"I didn't like that Anoush is treated as a neural network"* (P19). Both the author of this content and this reader self-identified as Armenian. Other Armenian learners did appreciate the same content though, highlighting how learner-made content covering sensitive cultural topics can be received very differently.

Beyond culturally-centered storytelling, scenes that introduced new characters out of the blue or that adopted a specific style of genre (e.g., fan-fiction, horror, etc.) tended to receive poorer reviews overall. For example, a scene titled *"Morgana explains a gradient,"* ventured into fan-fiction based on Knights of the Round Table lore, and it received the lowest emotional enjoyment rating of all platform content (mean: 2.33/6, std dev: 0.94). In post-study feedback from surveys and interviews, the majority of learners disliked these types of scenes because even when they contained *"fun elements"* (P1) or *"relevant AI content"* (P9), their tone felt out of place with the broader narrative. Such comments were evenly expressed by learners who identified as Armenian and non-Armenian.

At the same time, some learners very much enjoyed creating these styles of scenes and were not discouraged in doing so even if the scenes received little to no traffic on the platform. These findings again indicate the importance of designing narrative learnersourcing

experiences that balance community preferences with opportunities for individual creative freedom.

3.6.4 Gender Differences

Across a number of the quantitative metrics and qualitative feedback, gender seemed to play a role in learning outcomes, engagement levels, and overall user experience.

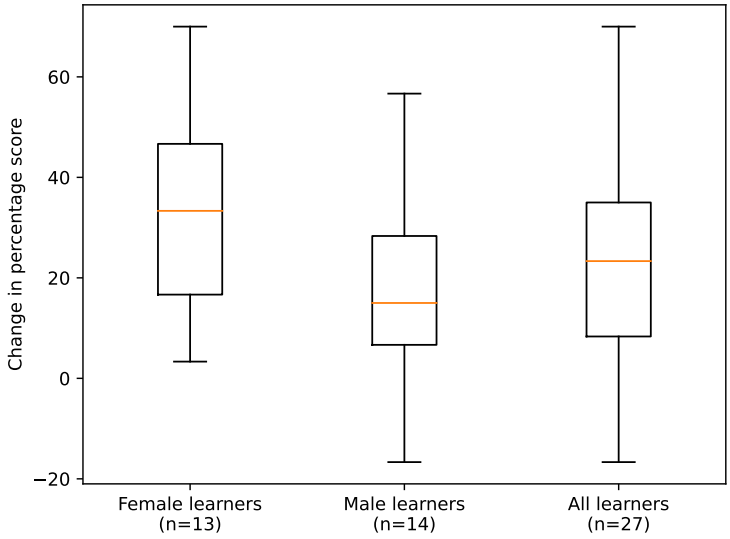


Figure 3.8: Female learners had a greater improvement in mean score from pre to post AI knowledge exam.

Table 3.5: Comparing female and male learners on assessed, self-reported, and logged metrics of learning and engagement.

Metric	**Measured by**	**F Mean**	**M Mean**	**p**	**d**
Baseline knowledge	% correct on pre-exam	36%	40%	.62	0.20
Baseline self-efficacy	Pre-study assessment [0,1]	0.53	0.51	.53	0.25
Knowledge gains	Pre to post % change	32%	17%	.05	0.81
Attentiveness	Time per word (sec)	2.8	0.02	.04	0.87
Engagement	Time spent (hr)	26.4	23.7	.88	0.06

First, a pattern was observed where female learners appear to have slightly higher learning gains compared to male counterparts. This difference is illustrated by Figure 3.8. To better understand these trends, female and male learners are analyzed in terms of various

objectively assessed, self-reported, and automatically logged measures of knowledge and engagement. Table 3.5 summarizes results.

No significant gender differences were seen in prior knowledge at baseline nor self-efficacy at baseline, as well as negligible effect sizes for each of these comparisons, suggesting that any observed differences in post-study assessments do not simply reflect these baseline levels.

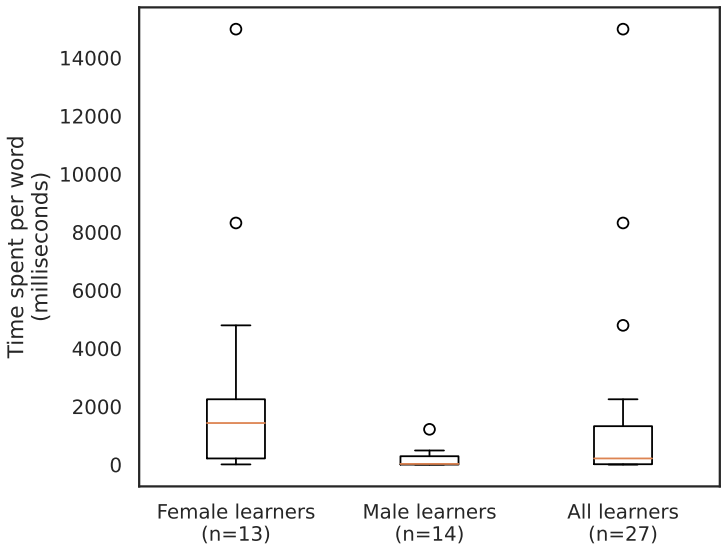


Figure 3.9: Female learners tended to have higher attentiveness scores compared to male learners.

Noticeably higher female attentiveness was observed, as illustrated in Figure 3.9. This may be due to male participants reading faster than female participants or male participants skimming some story content, although the literature on gender differences in text processing speed suggests the former interpretation is unlikely. Specifically, relevant work has shown that females outperform males in alphabet-related processing speed tasks as well as other reading and writing tasks [Roivainen, 2011]. It is also worth noting that the measure of engagement (time spent on the platform) shows a less pronounced gap between female and male learners (see Figure 3.10).

Based on qualitative analysis, many male learners did consider the narrative interesting, though non-essential to the learning process. Some even saw it as a distraction. This is not

to say these learners did not find the platform meaningful; on the contrary, UES-SF metrics for both genders were roughly comparable (see Figure 3.11). Rather, these findings serve as indicators that female and male learners used and benefited from the stories in different ways.

Specifically, female learners tended to self-identify with the story's main and supporting characters (nearly all of whom were also female) and explored AI concepts by focusing on aspects of characterization. On the other hand, male learners did not self-identify much with the characters, but they were nonetheless interested in the narrative and focused more on the provided metaphors to make sense of AI concepts. To help contextualize these general observations, two participants, P3 and P17, are presented as vignettes.

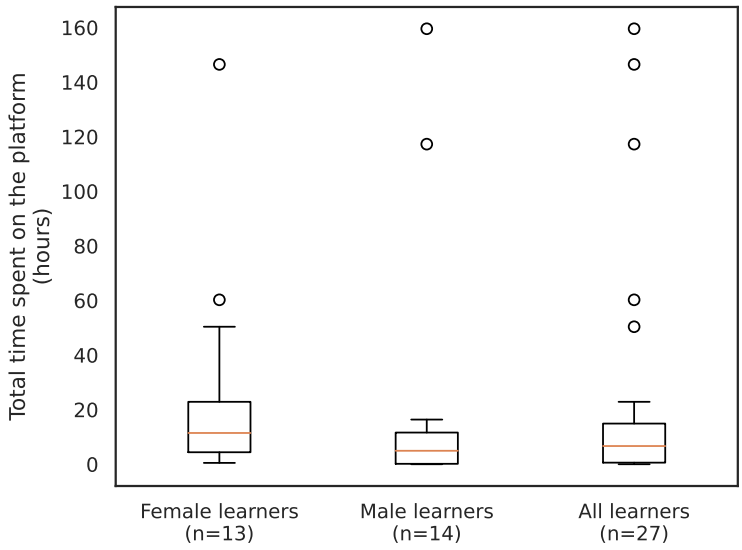


Figure 3.10: Engagement time on the platform was relatively similar for female and male learners.

P3 identified as female, was 17 years old, and lived in the US. She reported minimal computer science experience and no connection to Armenia. She learned about the study through a future college classmate. After joining the study and engaging with the platform, P3 met many other learners who, like herself, were considering a STEM degree. The system's community aspects were a major draw for her and kept her highly engaged. In her

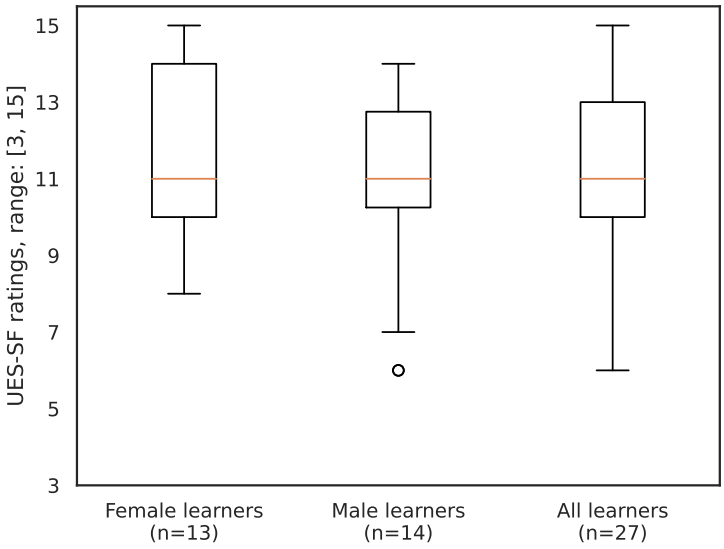


Figure 3.11: The user experience ratings by female and male learners were relatively similar.

words, *"Where I felt most strongly about community was seeing everyone else's code and thinking, 'Wow, this was created by multiple people, even if I didn't meet those people.' "*

P3 also noted that learning through stories and then creating stories herself made the learning process much more approachable and enjoyable. She also found pleasure in learning about a culture different from her own: *"I like [the cultural and storytelling] aspects a lot. At least from my point of view, you don't really hear a lot about Armenian culture and inside someone's view of it. And I love learning a lot about different kinds of cultures."*

These stories also made the AI topics more legible to P3: *"Being able to metaphorically make connections to the topics made learning the content much easier. I would be like, 'I don't get this,' but then in the next sentence, I would be relating to something that Carol [a story character] would be talking about—and I'd be like, 'Okay, I understand it!' The statues scene was my favorite, when Carol was admiring the statues, pointing out different aspects, and stating what they meant to her, beyond just being 'stone' or 'marble.' "* Further, P3's perspectives illustrated how narrative learnersourcing had a powerful impact on her attentiveness and self-efficacy, as stories felt gripping and relatable. This experience was representative of many other female participants' reactions to the platform.

P17 self-identified as male, was 18 years old, living in Greece, and self-reported significant prior computer science experience from extracurricular courses. At first, he described the stories as more suitable for younger students, calling them *"childish"* and *"a pointless thing to read."* At the same time, P17 very much enjoyed creating new story content for the platform. He found level 4 especially interesting: *"Level 4 was interesting because there were the ethical concerns, and it was more for real life situations and problems that I might face. It was a great reminder that we need to think about those issues. They are very important."* P17 said he did not consider the community aspects of the platform to be personally very important to him, though he believed others might find that more meaningful. His perspectives, while skewed slightly more negative, showed similarities to many other male learners, illustrating the need for systems like this to support story creation, exposure to broader AI considerations such as ethics, and other features that resonate more deeply with users like P17.

3.6.5 Platform Scalability

Much of the motivation in developing a learnersourcing platform is this approach's promise in promoting a virtuous cycle of engagement, contribution, and maintenance that can enable the community to sustain itself over time as it grows to reach and benefit increasingly broad groups of learners. Here I examine important scalability challenges, including the cold start problem (having sufficient content when the platform is relatively new), issues of content quality, and how new users can become aware of the platform through organic, learner-driven referrals.

3.6.5.1 Tackling the cold start problem through human-AI collaboration

When considering scalability, it is important to understand the early stages of a platform's lifecycle, including how much content needs to be pre-seeded when the platform is first activated in order to provide a sufficient initial foundation on which early members can

then build. For this study, the platform was seeded with learner-made content sourced through the three preliminary case studies.

A main difference between the case studies and final evaluation study was that my co-authors and I did a considerable amount of hand-holding of learners in the case studies (precisely to get around problems like writing stories without a lack of prior content to build on), but a fully hands-off approach was taken in the final study, letting learners act as facilitators for each other. The experience throughout these phases indicates that creators of new instantiations of a narrative-based learnersourcing platform like Apricot Stone City will need to work with early adopters to draft initial content and/or produce it themselves. Then once that ball is rolling, learners are enthusiastic about keeping the contributions coming.

To explicitly gauge such contributions made to the platform, the number of changes made by story architects was calculated (total number of scene contributions: 104, mean: 8, median: 6, min: 1, max: 20). The average word count per story contribution was 590, and the cumulative word count of all content on the platform was 88,636. These results represent a meaningful amount of work, particularly when considering that all story content was in English yet many of the users were English as an Additional Language (EAL) speakers.

This level of output was supported by learners leveraging ChatGPT as a writing collaborator, which permitted acceptable quality and more consistent content creation by learners who were mostly novices in both writing and AI. In interviews, learners appreciated this story consistency, while feeling it was important that stories were co-written by a human rather than simply automatically generated by large language models or other forms of generative AI. The use of such tools as part of collaborative human-AI writing activities thus may be a promising avenue to promote scalability and minimize cold start issues, while preserving the essential aspects of user-driven content creation that are inherent to learnersourcing as a pedagogical strategy.

3.6.5.2 Gauging quality of learnersourced content

A key issue in crowd-powered systems is ensuring the quality of user-generated content. To assess the quality of learnersourced content on the platform, an expert review of all learner-created story scenes was performed. Criteria for gauging correctness and helpfulness (both on a scale of 1–5) based on Denny et al. [2023]'s evaluation of learnersourced content were used, averaging independent scores per metric. For content correctness, a mean of 3.91, median of 4, and standard deviation of 1.03 were found. For content helpfulness, a mean of 3.38, median of 4, and standard deviation of 1.41 were found. These results indicate that a majority of learners were able to produce content on the narrative-based platform that is at least "satisfactory" in terms of teaching quality.

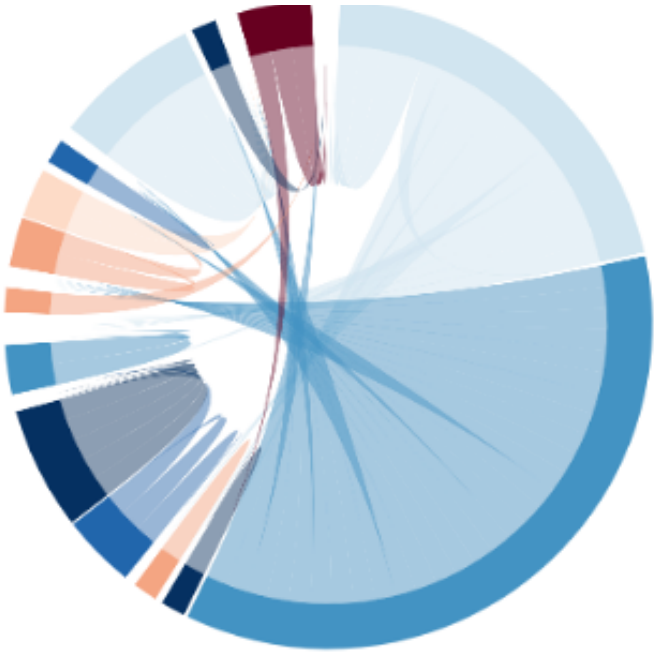


Figure 3.12: Referrals made and accepted by users on the learnersourcing platform.

3.6.5.3 How learner referrals help scale and sustain the platform

While only outcome data from the study cohort of N=27 recruited participants were analyzed, I was interested in how much organic growth the platform could generate in just one week through unprompted, learner-driven word-of-mouth. During the course of the

study, 105 users joined the platform and stayed active. (Users are considered "active" if they successfully create an account and interact with at least one scene a day for three days. A referral is considered "accepted" only if the recipient becomes an active user.)

Most of this userbase discovered the platform through referrals from friends or acquaintances. Figure 3.12 illustrates the flow of referrals accepted by new users. Each distinct arc on the circle's circumference denotes one user, with the total circumference representing all 105 users active on the platform. The distance along the circumference covered by any particular arc represents the number of referrals that particular user made. Connections between the arcs inside the circle depict which users accept which referrals.

Analyzing the referrals, new users tend to join in groups linked to a specific referrer, with two users (P21 and P25) driving over half of all accepted referrals. Identifying and studying these highly motivated referrers would be a valuable next step to understand strategies for reproducing this effect. Surprisingly, the main distinguishing factor of the largest "super referrer" was not prior knowledge or experience in computing but rather his enthusiasm and belief in the concept of narrative learnersourcing, which bodes well for its appeal factor and potential for uptake.

3.7 Discussion

To promote learning experiences for AI literacy that are inclusive, immersive, and personally meaningful, this research explores a strategy that combines learnersourcing with the intuitive, engaging benefits of narrative-based learning. Specifically, the Apricot Stone City platform leverages culturally-specific metaphors to wrap AI concepts in an interactive narrative adventure.

3.7.1 Contributions and Reflections

My co-authors and I followed an iterative, user-centered design approach spanning three formative studies with target users along with an extended deployment of the resulting system. These efforts also involved forming and maintaining a strong collaboration with the extracurricular program partner, TUMO, as part of connecting with learners to understand their needs, preferences, and experiences around novel learning environments.

The specific contributions of this work include:

- An innovative narrative-based learnersourcing approach that uses stories as a vehicle for active, collaborative learning of AI concepts.
- Instantiation of this design approach in a usable online platform, with features and functionality informed and refined through multiple rounds of learner-centered design.
- Rich findings from a deployment study that demonstrate narrative-based learnersourcing can promote knowledge gains, topic interest, and positive learner experiences, including for novice learners and those from backgrounds traditionally underrepresented in computing and STEM fields.

3.7.2 Community Partners and Culturally-Situated Narratives

This research was done in partnership with TUMO, which provided supporting resources, far-reaching access to study participants, and real-world extracurricular learning contexts. It would be worthwhile for future work to consider testing learnersourcing systems in partnerships with other types of organizations, such as smaller-scale extracurricular efforts, formal educational settings, and community-based organizations that engage with additional age ranges.

Given TUMO's roots in Armenia and the initial case studies in Armenian contexts, culturally-situated storytelling became an important foundation of the approach to

narrative-based learnersourcing. The design process evolved to center cultural expression as a key aspect of the learning experience, with the formative design phases exploring cultural and cross-cultural factors through in-depth, direct engagement with learners and their communities. These findings align with work showing narratives in online communities can be amplifiers of participation, shared values, fulfillment, and emotional connection [Leyton Escobar et al., 2014]. In engaging with such issues, this work spans sociology, pedagogy, and narratology. I hope this research inspires additional interdisciplinary scholarship at the intersections of HCI and broad humanistic and technical domains.

3.7.3 Narrative-Based Learnersourcing for Inclusive Learning Communities

While this chapter focused on the deep and extended human-centered design process to develop a narrative-based learnersourcing system and verify its preliminary effectiveness, an important next step is establishing the approach's scalability. The findings do indicate that users are able and motivated to learn from and build on prior learnersourced content.

Promise was also seen in the organic growth of the platform's userbase through word-of-mouth referrals. The emergent behavior observed in the referral patterns aligns with prior HCI research on "super users" as exceptionally influential [Whitten et al., 1986]. Such trends could also be an early indication of preferential attachment—a property where nodes in a graph acquire new links at a rate relative to the node's existing degree. Colloquially called the "rich get richer" effect, this mechanism has contributed to the growth of large social networks and crowdsourcing projects, including Wikipedia [Jeong et al., 2003, Capocci et al., 2006].

Taking a step back, I reflect on why scalability matters. In this study, scaling the platform resulted in not only more content quantity, but also higher content quality as more learners became facilitators. In turn, new and improved content kept learners engaged and motivated their outreach to new generations of users. Similarly, the sense of community

flourished and persisted across the four learner cohorts (the 3 case studies plus the evaluation study) based on post-study interviews as well as on-platform interactions (e.g., helping each other debug code, learn GitHub, or troubleshoot the development environment).

Lastly, by conducting the case studies and final study in different learning settings, it was observed that the platform could be effectively employed in a range of contexts. It would be desirable for future work to explore the value of narrative-based learnersourcing in a variety of other non-formal and formal learning environments [Johnson and Majewska, 2022].

3.7.4 Narrative as Empowerment Tool

In addition to the positive outcomes seen for learners across the study sample, I observed that narrative-based learnersourcing can be an effective means of promoting sustained engagement, AI mastery, and self-efficacy for young women specifically. Such findings are especially significant given women are traditionally underrepresented in computer science and strategies are needed to enhance girls' interest and retention in the field [Google, 2014, Google Inc. and Gallup Inc., 2016, Master et al., 2017].

This work additionally responds to "adult-centrism" issues raised by Prilleltensky et al., who claim that most work interprets the realities of young people from the point of view of an adult, thus depriving young people of power [Prilleltensky et al., 2001]. When considering Prilleltensky's terminology, narrative-learnersourcing could empower young people by extending access and control of the "key dimensions of power"—namely, access to valued resources, opportunities for participation and self-determination, and opportunities for the development of competence and self-efficacy.

3.7.5 Entering the Age of AI

Finally, I situate this work in relation to AI literacy education as a movement. Prominent advocates of AI literacy curricula are not only interested in responding to but also shaping

the trajectory of technological and social development. This position was articulated by Stefania Giannini, the UNESCO Assistant Director-General for Education, in her 2023 report on generative AI and the future of education:

> *"In our environment of AI acceleration and uncertainty, we need education systems that help our societies construct ideas about what AI is and should be, what we want to do with it, and where we want to construct guardrails and draw red lines. Too often we only ask how a new technology will change education. A more interesting question is: How will education shape our reception and steer the integration of new technology—both technology that is here today and technology that remains on the horizon?"* [Giannini, 2023].

The Apricot Stone City system was designed based on an understanding of the connectedness between these two questions. That is, learners are positioned as both consumers and creators of learning content, thereby supporting their ability to actively bridge gaps between these roles by employing conceptual and technical critical thinking and by leveraging the connective tissue of narratives.

3.7.6 Limitations and Future Work

In acknowledging limitations and associated opportunities for future work, it is first noted that learnersourcing systems have well-documented content quality challenges, such as inaccurate or unhelpful learner-made resources [Pirttinen et al., 2023]. The system's design choices aimed to promote quality, for instance through learner-driven content evaluations (e.g., post scene ratings) and control mechanisms (e.g., code reviews and bug reports to facilitators). However, evaluating the relative effectiveness of each of these mechanisms was not the primary focus of this chapter. Although the expert review indicated that learner-made content was generally correct and helpful (see Section 3.6.5.2), more work is needed to unpack the many design questions related to quality control.

Next, it is worth pointing out that the system's logging capabilities allowed reliable assessment of many on-platform behaviors; however, some sub-tasks involving content creation (e.g., writing code locally) were more difficult to track. During in-person case studies, direct observation could overcome this issue, but self-reported data had to be relied upon during the final evaluation study. Future development work could address such limitations by implementing more comprehensive tracking support to enhance scientific investigations, while staying sensitive to learner privacy, data security, and best practices around disabling tracking to avoid unnecessary surveillance of users.

Finally, a 1-week period was chosen for the evaluation study given that duration reflected the standard length of many of TUMO's extracurricular programs, plus I wanted to understand whether knowledge, attitudes, and user experiences could be meaningfully shaped in this relatively short timeframe. While promising outcomes were found along these lines, it is critical to undertake more extended studies to investigate novelty effects, sustained engagement, and ultimate scalability of narrative-based learnersourcing.

3.8 Conclusion

This chapter presented a pedagogically-grounded, narrative-based learnersourcing approach to teaching AI, with a focus on extracurricular contexts. Following an iterative, user-centered design process involving three preliminary case studies in multi-cultural settings, I instantiated my approach in a web-based platform named Apricot Stone City. Evaluating my system in a deployment study with N=27 participants, I find that users experienced a significant increase in AI knowledge, showed positive shifts in self-efficacy over the course of the study, and were motivated to engage and re-engage with increasingly advanced AI concepts. Measures of user experience indicated that the narrative elements were a main contributor to meaningful engagement on both cognitive and emotional levels, and participants expressed a strong sense of community on the platform. Wrapping AI

concepts into story content also made the learning experience more approachable for many users, particularly female learners and those with little prior experience with computer science. Based on my insights, I offer reflections on the value of community partners and culturally-sensitive narratives in creating inclusive learning experiences, and I encourage the HCI community to see narratives as an empowering educational tool as we enter the age of AI. In Chapter 4, I carry this view of narrative into a very different setting, examining how AI infrastructure might capture and circulate the personal knowledge that caregivers hold about the individuals they support.

Chapter 4

CareInsights: AI Infrastructure for Person-Centered Dementia Care

4.1 Introduction and Motivation

Despite growing awareness of the complex needs of people living with dementia (PLWD), care practices in long-term residential facilities often fall short of providing humane and responsive support. In many under-resourced settings in particular, caregivers rely on crisis-driven routines and reactive interventions to manage residents' behaviors, often by overusing antipsychotic medications to reduce agitation or resistance to care [Ballard et al., 2009]. At the same time, systemic understaffing and high turnover reduce opportunities for meaningful engagement, leading many residents to experience long periods of inactivity, emotional neglect, and social isolation [Heiks and Sabine, 2022]. These conditions lead to care environments where distress is medicalized, agency is minimized, and residents are often treated as passive recipients rather than individuals with unique histories, preferences, and emotional needs.

Person-centered care offers an alternative framework that emphasizes dignity, autonomy, and the centrality of the individual's story in all aspects of care. This approach encour-

ages caregivers to move beyond task-oriented routines and tailor interactions to each resident's identity, values, and lived experience [Kitwood, 1997, Brooker and Latham, 2015]. Person-centered care has been linked to improved emotional well-being, reduced behavioral symptoms, and enhanced quality of life for PLWD [Fazio et al., 2018]. In this work, I extend the ethos of person-centered care by examining how AI-enabled documentation infrastructure might represent health information in narrative form for regular caregiver support. This format mirrors informal caregiver-to-caregiver information-sharing practices, and is naturally well-suited to capture the needs and lived experiences of PLWD.

Within person-centered dementia care, *memory therapy* refers to a family of non-pharmacological interventions such as reminiscence, art and craft activities, and music-based therapy. These interventions stimulate autobiographical memory, preserve identity, and strengthen social bonds for PLWD [Fazio et al., 2018, Matziorinis and Koelsch, 2022]. For instance, *music-based memory therapy* draws on the relative preservation of neural circuits linking musical processing and emotional regulation in dementia [Clark and Warren, 2015] and uses familiar songs to elicit music-evoked autobiographical memories that improve mood and facilitate connection [Matziorinis and Koelsch, 2022]. Such forms of memory therapy and person-centered care depend on caregivers having access to rich, up-to-date knowledge about the individuals they support. Knowing a resident's preferred daily routine, emotional triggers, past occupations, or signs of discomfort can enable more empathetic and effective care. However, this situational awareness of the person behind the symptoms is often undermined by fragmented documentation and the overwhelming demands placed on caregiving staff. Without reliable mechanisms for capturing, sharing, and organizing this contextual information for caregivers, it will be difficult to bring the ideals of person-centered care into common practice.

Although electronic health record (EHR) systems are positioned as the backbone of modern care, their design priorities emphasize compliance, billing, and incident reporting, not relational care. As a result, they rarely accommodate the kinds of narrative, expe-

riential, and biographical information that caregivers rely on to respond empathetically in the moment. The very *infrastructure* intended to support care ends up sidelining the personal knowledge that makes it humane. Despite widespread recognition that quality person-centered care relies on contextual awareness, existing systems are ill-equipped to capture or elevate the insights most relevant to frontline caregivers [Anantapong et al., 2022]. In the absence of systems designed to capture these nuances, contextual knowledge circulates through fragile, informal channels such as brief hallway exchanges, handwritten shift-change notes, or personal recollections passed between staff. In environments marked by high turnover, burnout, and chronic understaffing, these insights are easily lost, eroding continuity of care and reinforcing reactive, task-driven routines [Heiks and Sabine, 2022].

Ethical frameworks of care remind us that attentiveness to the person behind the task is not merely a professional virtue but a moral responsibility [Tronto, 2020]. However, caregivers are seldom given the time or tools to uphold this standard. Manual documentation is burdensome and inconsistent, and the details that animate compassionate care (e.g. small gestures, emotional cues, moments of connection) often fade before they can be recorded. Over time, these losses compound, shaping organizational cultures that, despite individual caregivers' intentions, struggle to collectively value and operationalize person-centered care. I propose that closing this gap requires reimagining biographical and emotional context as meaningful clinical data and designing infrastructures that make such knowledge visible, shareable, and actionable in everyday practice.

Specifically, I explore the following research question:

> *How might we design AI-enabled technology that captures relational knowledge and reduces documentation burdens to enable more person-centered care in resource-constrained facilities?*

To undertake this work, my co-authors and I partnered with Memcara Inc.[1], an organization of certified dementia practitioners that trains and employs specialists in memory

[1] https://www.memcara.com/

therapy and operates across dozens of long-term care facilities in the United States and Canada. From seven of these facilities, we recruited $N = 19$ facility staff caregivers who contributed to ethnographic observations, collaborative system development, and in-situ evaluation activities.

The contributions of this work can be summarized as follows:

- First, I present a fieldwork-driven qualitative analysis based on observations and semi-structured interviews across seven long-term care facilities. This analysis identifies when and how relational knowledge breaks down within existing care workflows, revealing key opportunities for infrastructural support.

- Second, I introduce *CareInsights*, an AI-enabled documentation system designed and iteratively refined through close collaboration with caregivers. Integrated into facilities' existing EHR platforms and embedded within Memcara's tablet-based memory therapy application, CareInsights captures short, session-bounded recordings and event metadata. Machine-learning models detect salient behavioral and affective moments, transforming them into concise, time-stamped narrative snippets stored on facility servers. These narratives feed an embedded chat agent that uses retrieval-augmented generation to provide real-time, context-aware support. Caregivers can ask about a resident's life history, recent clinical developments, or stimuli that evoke positive or negative reactions in order to receive concise, traceable responses synthesized from passively captured session data and relevant health records. This person-centered approach to documentation preserves relational context while reducing staff burden and avoiding privacy-invasive continuous monitoring.

- Finally, I examine the infrastructural and ethical implications of embedding computational systems within high-stakes care contexts, arguing that narrative documentation provides a pathway for AI to enhance human caregiving relationships.

4.2 Background and Related Work

This section reviews foundations for this case study, including person-centered care theory, care ethics, sensemaking and documentation challenges in caregiving, infrastructure-oriented approaches to relational care, and specific therapeutic approaches in the context of dementia care.

4.2.1 Care Ethics and Person-Centered Practice

Modern efforts to improve dementia care have centered on supporting "personhood" by preserving the identity, dignity, and agency of people living with dementia even as their cognitive abilities wane. The foundation of this approach can be traced to Tom Kitwood's critique of biomedical dementia models in the late 1980s [Kitwood, 1989]. Drawing on Carl Rogers' humanistic psychology [Rogers, 1961], Kitwood reframed dementia not as a purely neurological deficit but as a condition profoundly shaped by social and relational context [Brooker, 2003, Dewing, 2008]. He introduced the concept of personhood as something sustained through relationships defined by recognition, respect, and trust rather than by cognition alone [Kitwood, 1997]. Later person-centered care frameworks sought to translate these values into practice.

Providing holistic care in long-term facilities is essential to improving quality of life for PLWD, including enhancing emotional well-being, social engagement, cognitive stimulation, and physical functioning [Williams et al., 2015]. *Person-centered care* goes beyond treating symptoms and emphasizes understanding a PLWD's history, preferences, and values to ensure compassionate and responsive care practices [Manthorpe and Samsi, 2016]. Person-centered care is crucial in the context of dementia because it upholds the dignity of individuals and addresses their distinct cognitive and emotional requirements [Fazio et al., 2018, World Health Organization, 2015]. This model improves interactions through person-centered communication [Williams et al., 2018], minimizes the dependence on po-

tentially harmful antipsychotic medications for managing agitation [Fossey et al., 2006], and significantly contributes to the overall well-being of both care recipients and caregivers [Brownie and Nancarrow, 2013].

Theories of care ethics illuminate what person-centered practice requires of systems that support it. Tronto's ethic of care [Tronto, 1998, 2020] emphasizes that attentiveness, responsibility, and responsiveness depend on what is recorded and acknowledged in the system. Pirolli and Card's sensemaking model [Pirolli and Card, 2005, Klein et al., 2007] describes how practitioners iteratively notice cues and interpret them through prior knowledge. Together, these frameworks highlight that care quality relies not only on the availability of information but on how it is integrated and made actionable across teams and shifts.

Care ecologies can be conceptualized as complex adaptive systems, offering a framework for comprehending the intricacies of care delivery, its inherent unpredictability, and the feedback mechanisms that influence the well-being of both caregivers and care recipients. This framing emphasizes the dynamic interactions among various constituents—including caregivers, care recipients, families, administrators, and any interactive technologies—and how these parties' behaviors are adaptive, emergent, and non-linear. These qualities mean that interventions, including digital agents, must similarly be adaptive to shifting contexts to accommodate the evolving nature of the environment and its users [Carroll et al., 2023, Chaffee and McNeill, 2007, Gilman, 2021, Pype et al., 2018]. Person-centered care guidelines call for engaging family and friends in shared decision-making and communication given their knowledge and longstanding bonds with the PLWD [Podgorski et al., 2021], and they position caregivers as advocates who foster ongoing relationships with residents' families to enable person-centered practice [Brownie and Nancarrow, 2013].

4.2.2 Sensemaking and Documentation in the Context of Caregiving

In practice, long-term care environments rarely provide the infrastructure needed to sustain collective sensemaking. Staffing shortages, high turnover, and time constraints often force caregivers to rely on memory or informal communication rather than consistent documentation [Heiks and Sabine, 2022, Martin et al., 2023]. Records tend to be fragmented, verbal, or non-standardized, leaving new or temporary staff without access to essential context about residents' routines, preferences, or behavioral patterns [Anantapong et al., 2022, Hermann et al., 2022].

While mainstream EHR systems promise continuity, they are primarily optimized for regulatory compliance and billing rather than relational care. Platforms such as PointClickCare[2] focus on structured clinical data while offering limited support for residents' histories, affective cues, or personal preferences [Anantapong et al., 2022]. Optional narrative fields are rarely populated due to limited staff time, inadequate training, and poor interoperability [Hermann et al., 2022]. As a result, critical relational knowledge—such as what soothes or engages a resident—often remains tacit, locked in the memory of individual caregivers, and lost when staff turnover occurs.

These breakdowns in documentation disrupt the sensemaking loop, forcing caregivers to start each shift with limited context, repeat ineffective interventions, and miss opportunities to prevent distress. Evidence across long-term care and hospital settings links poor handoffs to preventable incidents, while structured, narrative approaches improve continuity, intervention accuracy, and safety outcomes [Poelen et al., 2023, Starmer et al., 2014]. These findings underscore the need to design documentation infrastructure centering context—systems that make person-centered insight durable, interpretable, and actionable across shifts.

[2]PointClickCare is one of the largest providers for long-term care facilities [PointClickCare Technologies Inc., 2025]; as of 2024 it was used by "more than 27,000 long-term and post-acute care providers, 3,600 ambulatory clinics, 2,800 hospitals, 350 risk-bearing providers, 70 state and government agencies, and every major U.S. health plan" [PointClickCare Technologies Inc., 2024].

4.2.3 From Documentation to Sensemaking Infrastructure

The breakdowns in documentation and sensemaking described above highlight the need for infrastructure that can connect fragmented records, interpret relational context, and distribute it across teams in ways that fit care practices. Advancing intelligent, context-aware caregiving requires systems that integrate diverse data sources, support collaborative interpretation, and scale across care settings. Prior work in ubiquitous and collaborative computing has explored AI systems that aim to reduce caregiver burden through sensing and communication across clinical and home environments [Sharma et al., 2021]. Yet many remain fragmented, difficult to embed into daily practice [Hekler et al., 2016], and limited by persistent concerns around transparency, trust, and algorithmic overreach [Lukkien et al., 2024].

Recent human-centered design research points to narrative documentation as a promising approach for bridging these gaps. Embedding brief, biographical vignettes within clinical records allows caregivers to retain residents' personhood and respond with greater empathy and accuracy [Cherian et al., 2024, Dellkvist et al., 2024, Gui et al., 2023]. Narrative documentation, in this sense, becomes an infrastructural layer that sustains relational continuity in data-intensive care environments.

Such infrastructure must be co-designed to fit the routines, social dynamics, and privacy expectations of care work [Bhat et al., 2021]. In dementia care, this means moving beyond deficit-oriented models toward tools that preserve autonomy, identity, and connection [Brookman et al., 2023]. It also requires evaluating systems not only for technical performance, but for how they expand caregivers' practical capabilities and emotional bandwidth [Sen, 1999, Khullar et al., 2025, Khan et al., 2017, Goerss et al., 2024].

In long-term care, this perspective underscores the value of AI systems that synthesize insights from passive sensing, EHR data, and staff observations to enhance—rather than replace—relational work [Yang et al., 2019, Waddell et al., 2024]. For example, Das Swain et al. [2025] show that large language model–based "empathetic coworkers" can support

frontline staff's emotional labor without displacing human empathy. Such designs highlight AI's potential to mediate between data and lived experience, transforming fragmented observations into shared narratives that sustain continuity of care.

4.2.4 Person-centered Dementia Care

This subsection examines specific therapeutic approaches used in dementia care settings. Ridder et al. [2023] developed the Person-Attuned Musical Interactions (PAMI) model as a structured approach to therapeutic music engagement. PAMI sessions pair individualized music selections with narrative prompts, gentle singing, and tactile interaction, with the aim of eliciting reminiscence and emotional connection. In general, music-based memory therapy exemplifies how person-centered values can be enacted through technology-assisted practice. This has been demonstrated, for instance, by recent HCI work on mobile applications that support music therapy for family care settings [Vidas et al., 2024]. Studies show that, when used effectively, familiar songs tied to meaningful autobiographical memories can reduce agitation, elevate mood, and prompt moments of verbal clarity in people living with advanced dementia [Ridder et al., 2013, Moreno-Morales et al., 2020, Li et al., 2019, Jacobsen et al., 2015, McDermott et al., 2013]. These effects are linked to the relative preservation of neural pathways involved in music perception and emotional processing, which remain accessible even as other cognitive functions decline [Cuddy et al., 2017]. When particular musical stimuli are connected to salient emotions, they can trigger music-evoked autobiographical memories (MEAMs), temporarily restoring a sense of identity and enabling social connection. Even when residents do not recall sessions afterward, these interactions can have lasting effects—stabilizing mood and behavior disorders, reducing the need for medication, and nurturing more trusting relationships between caregivers and residents [Subramaniam and Woods, 2012, van der Steen et al., 2025].

The NIH Music-Based Intervention Toolkit [Edwards et al., 2023] provides a nationally recognized framework for evaluating and delivering music-based care, defining core

domains, measurement strategies, and reporting standards for both research and practice. These guidelines now shape how music interventions are designed and assessed in dementia care across the United States. Memcara adopts the NIH Toolkit as the basis for its own approach, embedding its measures into staff training, fidelity checklists, and routine care delivery.

Despite widespread endorsement of person-centered care principles, implementation in long-term care facilities remains challenging. Staffing shortages, high turnover, and time constraints limit caregivers' ability to tailor care or access resident-specific information [Heiks and Sabine, 2022, Martin et al., 2023]. Documentation practices are often fragmented, verbal, or non-standardized, meaning that essential context about a resident's routines or behavioral patterns is frequently inaccessible to new or temporary staff [Anantapong et al., 2022, Hermann et al., 2022]. These gaps limit caregivers' ability to make informed decisions and create relational moments of care, which are central to the person-centered care model. Research in care ethics and complex systems has called for better tools to support shared understanding across care teams and reduce cognitive load in high-pressure environments [Gilman, 2021, Pype et al., 2018]. Similarly, Ming et al. [2025] found that home healthcare aides benefit from data-driven insights to advocate for patient needs, underlining the broader need for care infrastructures that support frontline care work across settings. Their study highlights that while data can empower individual aides, its true power lies in aggregated patterns; workers and advocates must balance data comprehensiveness, privacy, and the burden placed on workers. This work takes up this call by focusing not just on caregiver tools or interfaces, but on the underlying data infrastructure that makes consistent, accessible, and contextualized information possible.

Building on these foundations, this work introduces a documentation infrastructure that links everyday therapeutic practice with data systems, showing how computational tools can support the situated, interpretive work of caregiving.

4.3 Approach and Formative Work

Building on the gaps identified in prior work, my co-authors and I sought to develop and evaluate a documentation infrastructure that embeds narrative sensemaking into the everyday realities of dementia care. The goal was to understand how computational tools could strengthen, rather than disrupt, the social and interpretive practices that make person-centered care possible. To do so, I adopted a human-centered, field-driven approach grounded in long-term collaboration with caregiving staff and organizational partners.

Long-term care facilities present a particularly demanding context for innovation. Chronic understaffing, high turnover, and fragmented documentation routines place constant pressure on caregivers' time and attention, leaving little margin for technologies that add complexity or fail to integrate with existing workflows. These conditions therefore serve as both a design constraint and a rigorous testbed for evaluating scalable interventions.

As shown in Figure 4.1, the approach unfolded across iterative stages that combined qualitative inquiry, technical development, and participatory evaluation. I first conducted in-situ observations and recordings of music-therapy encounters to capture the rhythms and relational dynamics of care. I then worked with clinical experts to co-create an annotated gold-standard dataset that informed model development. Finally, I implemented and iteratively refined an embedded intervention that integrates with caregivers' daily routines, enabling the capture, synthesis, and retrieval of person-centered insights at the point of care.

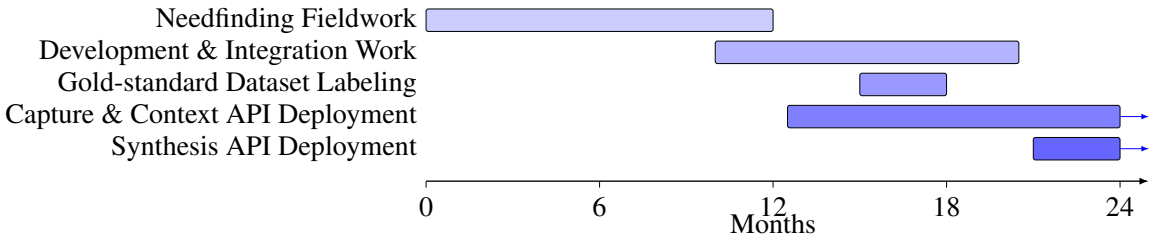


Figure 4.1: A timeline of the CareInsights study. The blue arrows pointing to the right indicate that CareInsights remains actively deployed.

4.3.1 Clinical Partnerships, IRB, and Participant Recruitment

My co-authors and I partnered with Memcara, a North American organization that delivers music-based therapy in collaboration with Universal Music Group through a custom tablet platform. As described in Section 4.2.4, Memcara aligns its program with the NIH Music-Based Intervention Toolkit [Edwards et al., 2023], which provides standardized measures for emotional regulation and social engagement. These principles guided both the design and evaluation of the system across multiple care settings.

Through Memcara's clinical network, we identified seven long-term care facilities (F1–F7) across the east coast of the United States and Canada. Throughout this chapter and its appendix, participants are identified by facility and role, so that F2-CG3 denotes the third caregiver (CG) at facility two and F1-CR6 the sixth care recipient (CR) at facility one (see Appendix B.6 for care recipient details). We selected resource-constrained, general[3] long-term care facilities to examine how documentation technology might support staff with limited dementia training. Although varied in size and setting, these sites shared common challenges of high turnover, limited staffing, and fragmented communication [Anesi and Kerlin, 2021]. All facilities used PointClickCare [PointClickCare Technologies Inc., 2025] as their primary EHR system and employed Memcara's tablet-based memory therapy platform, which delivers personalized playlists and real-time prompts.

Music-based memory therapy provided a structured and person-centered foundation for examining how documentation systems can capture and communicate relational insight. Memcara's sessions typically last 20–30 minutes and involve receptive listening, singing, prompt-based conversations, and movement—all activities that are intended to evoke observable indicators of engagement such as smiling, tapping, reminiscence, or emotional change [McDermott et al., 2013, Sousa et al., 2021, Moreno-Morales et al., 2020, Baird et al., 2018]. The therapy format supports systematic observation while minimizing intrusion, offering a clear lens into person-centered care in everyday practice. Crucially,

[3]Not dedicated memory-care facilities.

each session's impact depends on personalization: although the session structure is consistent, effectiveness arises from tailoring conversational prompts, visual stimuli, and musical choices to each resident's memories, preferences, and emotional associations. While Memcara's existing platform supported the delivery of personalized music, it did not capture or share the relational context underlying these choices. To fill this gap, I developed CareInsights, which was embedded into Memcara's existing platform to support documenting, organizing, and surfacing person-centered care insights that emerge during therapy sessions.

The study protocol was reviewed and approved by the Dartmouth College IRB. The review addressed ethical considerations related to working with people living with dementia, multi-site collaboration, and data handling across institutional boundaries. Facility administrators led recruitment, contacting families and obtaining consent from the legally authorized representatives on record when residents could not consent directly, following the determinations of capacity to consent that the facilities are required to maintain under their own regulatory obligations. I prioritized inclusion of individuals with advanced cognitive impairment, following inclusive design principles [Holmes, 2020]. Cognitive status was reassessed by Memcara staff at enrollment using the Montreal Cognitive Assessment (MoCA) [Julayanont and Nasreddine, 2017]. MoCA scores were used descriptively in the study to contextualize participation rather than as an exclusion criterion.

4.3.2 Needfinding: Understanding Caregiving as Information Work

I began with ethnographic fieldwork and interviews across all seven sites to examine when and why care was breaking down. All of the facilities served mixed populations and were not dementia-specific, which meant that most caregivers lacked specialized training in dementia care. This created additional challenges in supporting residents with dementia, who gradually lose the ability to advocate for themselves, even as their needs remain highly personalized and often grow more sensitive. While emotional and physical labor were

common sources of strain, caregivers consistently described a distinct information burden: the difficulty of accessing or sharing key contextual knowledge. These details, such as redirection strategies, trauma-informed triggers or comfort routines, are critical for delivering responsive, person-centered care. When such information was missing (e.g., a preference for being spoken to from one side due to hearing loss, a past trauma that made physical touch distressing, or a childhood song that helped de-escalate agitation) the result was often avoidable distress, increased medication use, and emotional burnout among staff.

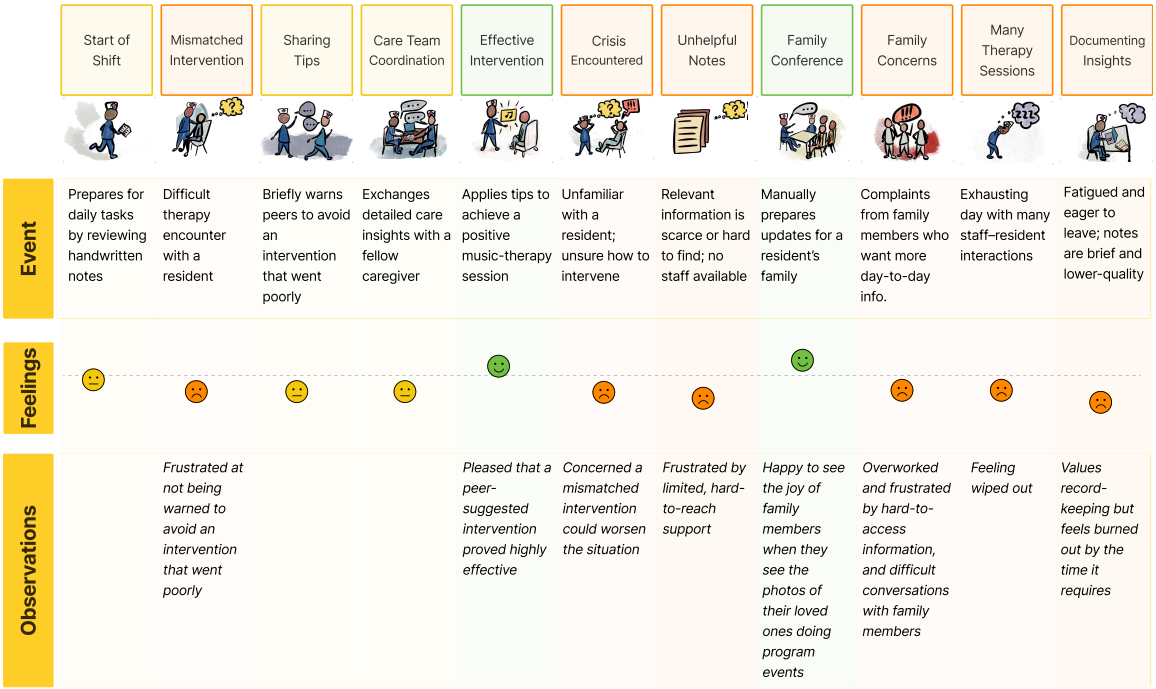


Figure 4.2: Annotated timeline of a caregiver's day showing common information-related tasks, emotional responses, and observed breakdowns across shifts. Illustrations were hand-drawn by the first author based on field observations. Notably, caregivers share short, memorable stories about care recipients as a form of peer-to-peer support; these stories combine background details with direct guidance and are adapted to the needs of each situation. While helpful, this channel is unreliable due to time pressure, poor record-keeping, and staff turnover. **Without a system to capture, contextualize, and synthesize these small but meaningful stories, their insights tend to remain isolated rather than guiding day-to-day decisions.**

As I analyzed these stories, I recognized that caregiving in these settings involved intensive, collaborative *information work*. Staff were constantly observing, interpreting, remem-

bering, and coordinating around tacit knowledge, yet they lacked systems to support that work. Mapping caregiver routines (for instance, see Figure 4.2) revealed recurring breakdowns in how information was captured, enriched with context, shared across shifts, and acted upon. These lapses often resulted in missed opportunities for personalized care, such as knowing that a resident was soothed by speaking about a grandchild or was triggered by references to a particular relative.

In response, many caregivers devised informal strategies to mitigate these gaps. These included hallway handoffs, working with fragmented EHR systems and "care cheat sheets"—simple, ad hoc documents used to track resident preferences and communication tips. When actively maintained, such tools were helpful to preserve context across shifts and support new or temporary staff. However, in practice they were inconsistently updated, difficult to access at the point of care, and raised concerns about privacy, up-keep, and long-term accuracy.

To better understand these breakdowns, I turned to the concept of *sensemaking*, which describes how individuals gather, interpret, and act on fragmented information in complex environments. Frameworks such as Pirolli and Card's sensemaking loop [Pirolli and Card, 2005] offer a useful model for how people process uncertainty through iterative cycles of information foraging and synthesis. While this framework shares some conceptual similarities with the observations from fieldwork, it does not fully reflect the situated and sequential nature of dementia caregiving. The findings revealed a progression of unmet information needs that consistently unfolded across four stages: **information capture**, **context enrichment**, **synthesis** and **point-of-care guidance**. These stages emerged directly from the realities of caregiving practice and became a practical foundation for the system design.

4.3.3 Baseline Gaps and Literature Guidance

Benchmark datasets offer important reference points for guiding the design and evaluation of models to assess a person's level of engagement. Initially, I considered existing

datasets such as DementiaBank [Lanzi et al., 2023], AffectNet [Mollahosseini et al., 2019], and recordings from the I-CARE Project [Schultz et al., 2018]. However, these datasets presented mismatches with the care context. DementiaBank primarily focuses on structured cognitive tasks, while AffectNet predominantly depicts younger, neurotypical adults and under-represents geriatric facial expressivity. The I-CARE corpus was thematically closer to the use case, as Steinert et al. [2020] trained and evaluated an emotion-recognition model for people living with dementia on 41 of its videos; however, the contextual setting and interaction dynamics differed. Consequently, I decided to curate a purpose-built "gold-standard dataset" tailored to music-therapy interactions in long-term care. I drew inspiration from Steinert et al.'s tablet setup because it provides stable framing, minimal disruption, and consistent video-audio quality. Although developed independently (see Section 4.3.6), the labeling schema overlaps with the multidimensional structure of the Video Coding–Incorporating Observed Emotion (VC-IOE) protocol used by Steinert et al. [2020], confirming that the measures sit squarely within the established scope of engagement assessment [Jones et al., 2015, Lawton et al., 1999].

Similarly, establishing an appropriate baseline model proved challenging. Off-the-shelf facial analysis tools, including the `deepface` library [Serengil and Ozpinar, 2021] and MediaPipe [Lugaresi et al., 2019], exhibited frequent misclassifications when initially applied to video recordings of memory therapy. Subtle but meaningful cues, such as quiet attentiveness or gentle emotional expressions, were consistently misinterpreted, which aligns with documented performance limitations of mainstream emotion-recognition models on older and cognitively impaired populations [Park et al., 2021, Kamikubo et al., 2022]. General-purpose meeting analysis tools such as Fathom [Fathom Video, 2024] and Fireflies.ai [Fireflies.ai, 2024] similarly produced irrelevant or misleading summaries, emphasizing their inability to handle the nuanced emotional and behavioral data that is critical in dementia care.

The difficulty in identifying publicly available baselines in AI, HCI, and healthcare literature highlights a significant gap. Publicly accessible datasets and models tend to be generalized or task-specific in ways that do not align well with the specialized needs of dementia caregiving contexts. Much of the existing research focuses on either diagnostic applications (e.g., early detection/classification of dementia [Wang et al., 2024b, Torabgar et al., 2025, Grueso and Viejo-Sobera, 2021]) or isolated recognition tasks involving healthy populations [Gupta et al., 2016, Dhall et al., 2018, Marquez-Carpintero et al., 2025], such as standard facial expression or emotional state classification [Gaya-Morey et al., 2026, Labzour et al., 2023, Alberto et al., 2023, Norori et al., 2021]. This is in part because regulatory and ethical constraints around capacity to consent and data governance make clinical video from dementia care difficult to capture and share. This limitation is discussed both in dementia-specific governance work [Milne and Brayne, 2020] and in healthcare-wide reviews of AI implementation barriers [Ahmed et al., 2023]. This general gap is made especially clear in Jiang et al.'s 2024 review, which highlights how underexplored facial expression analysis remains in the context of dementia. They argue that, when paired with other modalities, it holds strong potential for both detection and ongoing monitoring [Jiang et al., 2024].

Given these limitations, I developed a tailored, annotated dataset reflecting the unique multimodal interactions characteristic of dementia care. For a baseline comparison, I reference previous models trained for sufficiently similar tasks, where strong classification performance typically achieves F1 scores between 0.8 and 0.9 [Steinert et al., 2020, Torabgar et al., 2025, Vedernikov et al., 2024, Ahmed et al., 2025, Badawi et al., 2025]. For evaluating qualitative, narrative-based care descriptions, I rely on human judgment via Likert scales, benchmarking the system's outputs against expert-generated notes [Palm et al., 2025, Ganzinger et al., 2025].

4.3.4 Partnering with Caregivers to Translate Field Insights

The challenges uncovered in the fieldwork reflected a recurring cycle of breakdowns in how information was captured, interpreted, and applied. As summarized in Table 4.1, these breakdowns, which ranged from incomplete observations to a lack of real-time support, contributed to missed care opportunities, resident distress, and mounting caregiver strain. The fieldwork showed that caregivers continuously engaged in sensemaking—interpreting behavior and adjusting care in real time—but were hindered by fragmented systems and inconsistent documentation. The issue was not caregiver skill; rather, it was the lack of infrastructure to support their reasoning and workflows. Addressing this gap required more than improved documentation; it called for tools that align with caregivers' sensemaking processes.

Table 4.1: Breakdowns in documentation and sensemaking create reinforcing cycles of poor care quality and caregiver burnout.

Sensemaking Stage	**Caregiving and Documentation Challenge**	**Consequence for Care and Burnout**
Information Capture	Overworked caregivers struggle to record meaningful observations; documentation is minimized to compliance tasks.	Critical resident information is omitted or informally shared, leading to knowledge loss across shifts and missed opportunities for early intervention.
Context Enrichment	Fragmented, outdated, or inaccessible records make it difficult to retrieve context. Staff rely on memory or ad hoc communication.	Gaps in information continuity cause misinterpretations and reactive care, increasing emotional labor and frustration among staff.
Synthesis	Cognitive overload from irrelevant or low-value data; lack of tools to recognize patterns or build longitudinal understanding.	Staff experience decision fatigue and reduced situational awareness, leading to repeated mistakes and escalating behavioral issues.
Point-of-Care Guidance	Staff feel unsupported at the point of care, lacking real-time, resident-specific guidance.	Care becomes impersonal and efficiency-driven, losing attunement to residents' lived experiences; unmet resident needs lead to emotional distress and mounting strain across families, staff, and administrators.

The collaborative development process emerged from relationships established during a year of on-site fieldwork. Caregivers and activity staff participated as domain experts in three roles: (1) defining what counts as a "meaningful moment" during care work, (2) annotating initial sessions to create ground-truth labels, and (3) providing iterative, in-situ feedback on summaries and query responses embedded in Memcara's platform. This process aligns more closely with participatory annotation and human-in-the-loop supervision than with full co-design of the system's core architecture.

Sustained presence across the facilities built trust with aides, nurses, family members, activity staff, and administrators, creating the conditions for collaboration beyond observation. Further, I met weekly with Memcara, which served as a conduit between the research team and care staff by embedding early prototypes of CareInsights within their existing music-therapy platform. This integration allowed us to test features (e.g., session recording tools and contextual insight prompts) within workflows that staff already used. Memcara also supported implementation by coordinating deployments and helping us meet its standards for data security and platform stability.

Once recording functionality was in place, I observed that recording integrated smoothly into caregivers' routines without adding burden. Caregivers and families reported that the automated, context-specific nature of recording, which was limited to music therapy sessions, made it feel unobtrusive and minimized concerns about capturing sensitive or inappropriate moments. Design efforts then shifted toward analyzing recorded data and generating useful insights. With Memcara's support, I worked with caregivers to create the gold-standard dataset, which helped align the models with their interpretive practices. Once the Capture API (see Section 4.4) could reliably replicate this baseline, the system output was shaped using caregiver-provided notes and iterative feedback. Working closely with caregivers gave us confidence that the intervention would minimize disruptions to care and provide meaningful, person-centered support.

Given the focus on developing a generalizable intelligence layer, I intentionally limited front-end prototyping. Building a new interface without disrupting already fragile care routines would have required a more extensive co-design process. Instead, I opted for minimal modifications to Memcara's existing platform. As discussed in Section 4.8, deeper UI innovation remains an area for future work. In the subsections that follow, I trace the design process chronologically. Field insights about sensemaking practices shaped design decisions and guided the development of a labeled dataset for evaluation, culminating in the system described in Section 4.4.

Although all sites (F1–F7) used PointClickCare and Memcara's platform, their infrastructure varied: some had robust Wi-Fi and multiple dedicated tablets, while others relied on shared devices with intermittent connectivity. Staff also had differing levels of familiarity with digital tools, ranging from experienced activity directors comfortable with tablets to aides who rarely interacted with computers. These variations informed the decision to integrate CareInsights into the existing platform rather than introduce new hardware or logins.

4.3.5 Passive Video Capture via Existing Therapeutic Platforms

To avoid introducing new surveillance hardware or increasing caregiver workload, I leveraged Memcara's platform CareSuite (a tablet-based therapeutic application) already embedded in daily care routines. During scheduled therapy sessions, caregivers position the tablet to deliver digital engagement content, creating a natural opportunity to passively record caregiver–resident interactions. This positioning ensures consistent camera framing focused on active engagement, overcoming limitations observed in prior work with static or environmental cameras, such as occlusion, poor angles, and non-engagement. The session-bounded, tablet-based approach also mitigates privacy concerns relative to ambient cameras because recording is limited to scheduled therapeutic sessions: the camera is resident-facing and under therapist control, and the framing is intended to exclude common areas and bystanders. In practice, the FITL deployment surfaced footage with other residents in view (see Section 5.4.3.3). Video recording functionality was added to the tablet app, so that video could be captured automatically when a therapeutic session with a study participant begins, ensuring that caregivers do not need to take any additional action. Data storage and access are managed securely through the existing back-end infrastructure, with session identifiers linked to resident profiles to maintain longitudinal continuity.

4.3.6 Label Schema Development Process

The labeling schema was designed to mirror how highly experienced staff interpret resident behavior during music-based therapy. Rather than starting with a fixed taxonomy, I co-developed the labels with Memcara staff and facility staff caregivers, focusing on what they consistently called "Meaningful Moments"—short stretches of a session that felt worth documenting or sharing with families.

Together with caregivers, I analyzed 34 session recordings, most of which had minimal or no accompanying documentation in the EHR system or paper notes. To reconstruct caregivers' reasoning, the caregivers from the recordings rewatched the videos with us and first wrote retrospective care notes—i.e., what they would have entered had they been charting in real time. These were consolidated into two to three sentence narrative snippets that constituted Meaningful Moments.

Next, caregivers explained how they identified these moments by combining conversational content, observable behavior, and resident context. To formalize this judgment, they shaded blank timelines for each session to indicate changes across engagement dimensions. These visual timelines showed when and how sessions became meaningful and provided a second-by-second view of shifts in attention, emotion, memory, and behavior. The resulting annotations defined the engagement factors the model later learned to predict.

For additional detail, Appendix B.1 includes two example vignettes[4] that illustrate how caregivers marked and interpreted Meaningful Moments.

Aligning fine-grained engagement annotations with reflective narratives (i.e. Meaningful Moments) grounded the labeling schema in caregivers' own language and enabled a two-step, interpretable modeling pipeline (described in Section 4.5.1) that first learns the cues they attend to and then generates Meaningful Moments following their own logic and writing style.

[4]While these examples were manually generated by experts, the study results (see Section 4.7) indicate that CareInsights reliably generates comparable outputs at scale.

Through this process, five core engagement factors[5] were identified:

- Attention level
- Emotional expression
- Memory recall
- Reactions to therapy activities
- Behavioral symptoms associated with dementia (e.g., disturbance, confusion, or withdrawal)

These factors closely resemble those proposed in prior work on music-based memory therapy [Tan et al., 2019] as well as the NIH Music Toolkit [Edwards et al., 2023]. I observed that the caregiver-centered design worked well to distill abstract or research-focused measures into a practice-grounded rubric and a catalog of example Meaningful Moments that caregivers found interpretable and clinically actionable. The process started with Likert-scale annotations for each engagement factor to capture nuance, but caregivers found these too difficult to apply consistently, especially when trying to distinguish subtle attentional shifts for residents with more advanced dementia. Over time, I simplified to binary indicators and worked with annotators to set per-resident thresholds[6].

This labeling schema offers two key benefits. First, it grounds the training data in the same reasoning caregivers already use. Second, it enables a step-by-step modeling pipeline that mirrors how staff interpret sessions in practice. The system predicts these five factors on a per-second basis. Then, using simple rules, it detects spikes or shifts, such as a sudden onset of memory or emotion, and marks them as potential Meaningful Moments.

[5]Each factor is implemented as an array of one or more binary indicator variables (e.g., emotional expression is subdivided by emotion type). For clarity, these labels are presented here as five overarching factors to avoid redundancy.

[6]I began with Likert ratings to capture variations in dimensions like attention and memory (e.g., distinguishing recent vs. distant memories). However, caregivers found these distinctions difficult to apply reliably, particularly for attention. The transition to binary indicators, applying individualized thresholds per resident, better reflected what caregivers could feasibly and consistently determine in practice.

This design avoids the opacity of black-box classification and instead offers an interpretable path from raw behavior to notes caregivers might actually write. It also set the stage for the modeling pipeline to support evaluation and iteration with the same care teams who helped define the labels in the first place.

4.3.7 Early Feasibility Testing of Event Generation

Before implementing the full API pipelines, I developed an early-stage prototype to test whether Meaningful Moment generation could be feasibly achieved by splitting the task into two modular components: (1) identifying when a relevant event has occurred based on engagement signals and (2) generating a narrative summary of that event using the surrounding transcript and context. This modularity gave us confidence that each step could be developed and validated independently, rather than requiring a fully end-to-end system from the outset.

In the prototype, I started with narrative annotations that caregivers had created during the reflection on care notes (see Section 4.3.6), where they explained any documented Meaningful Moments. These caregiver-authored segments served as the gold standard. To better understand how these moments related to changes in engagement, I aligned each one with the underlying second-by-second annotations of engagement factors.

By examining how shifts in the continuous data corresponded to different types of events (e.g., a rise in memory recall before a personal story) I identified consistent patterns across multiple examples. From these patterns, I reverse-engineered a set of heuristics to detect event types based on change points in the engagement signals. Each change-point type was then paired with a GPT-4o prompt template, where key contextual details from the transcript and engagement data were used to fill in structured placeholders. This approach allowed us to generate summary descriptions that closely mirrored the kind of notes caregivers were already writing. As an example, I include the following prompt used for a change detected in the memory recall factor.

Prompt template for a Memory Recall Moment

You are assisting in analyzing recorded therapy sessions with older adults experiencing memory loss. Your task is to generate a clear and concise summary of a Meaningful Moment that occurred during the session. In this case, the resident appears to recall a personal memory. The goal is to produce a 2–3 sentence description that would be helpful to a caregiver or family member reading a session summary.
Use the following information to guide your summary:

- Transcript segment: [INSERT TRANSCRIPT HERE]
- EHR data (summary of key info): [INSERT EHR SUMMARY HERE]
- Events (e.g., songs played, images shown, discussion prompts) logged from Memcara's platform during the time window: [INSERT STIMULUS LOGS]

Output format: Return 2–3 sentences. Ensure your response is clear, emotionally sensitive and avoids speculation. Emphasize what the resident recalled and how the moment was expressed or triggered.

This approach allowed us to test whether GPT-4o could reliably produce summaries that were as useful and accurate as the notes caregivers typically wrote. After a short period of iteration and prompt tuning, I conducted a review with caregivers who had contributed the original annotations. They rated the generated summaries as meeting or exceeding the quality of their own documentation in terms of both utility and accuracy, particularly for family communication and session reflection. This validation helped confirm the feasibility of the layered approach: rather than attempting to generate Meaningful Moments directly from raw video, the system could first detect lower-level engagement features, apply a rules-based model to identify events and then generate summaries through templated prompts. Given that the synthesis layer already performed to caregiver standards when provided with accurate engagement labels, the remaining technical challenge was to automatically infer those labels from video, a task addressed by the ensemble modeling pipeline described in the following sections.

4.3.8 Gold-standard Dataset Creation

Having established the need for domain-specific supervision, as well as the desired outputs, I next focused on building a gold-standard labeled dataset tailored to the care setting. I began by asking caregivers to collect video recordings on participating care recipients using the embedded capture tool on the Memcara platform. After accumulating approximately a week's worth of footage, I piloted annotation strategies to generate engagement labels. However, early attempts to use external tools such as Labellerr [Labellerr, 2024] quickly surfaced major barriers. Caregivers were concerned about uploading sensitive videos of residents to third-party cloud servers, even with assurances of privacy protection. I explored whether de-identification strategies, such as blurring faces or removing audio, might address these concerns, but found that critical engagement cues were lost in the process. Preserving the context needed for accurate labeling inherently made full anonymization infeasible.

Beyond privacy, I found that existing commercial platforms designed for general-purpose annotation tasks included far more features than caregivers needed or wanted. Staff had very limited tolerance for learning new tools, particularly ones that required navigating complex interfaces or creating additional accounts. They emphasized that any labeling workflow needed to feel natural, efficient, and be integrated with their existing EHR tools, ideally letting them watch a video once at normal speed without pausing. This constraint made real-time, low-burden labeling essential. Together, these challenges reinforced the need to develop a lightweight annotation tool tightly aligned with caregiving practices, and to keep all data securely within institutional infrastructure.

To address these challenges, I developed a lightweight, real-time annotation tool, iteratively refined through caregiver feedback. The interface was modeled after familiar audio mixing timelines, with each engagement factor represented as a separate parallel track. Caregivers could toggle channels on or off while watching videos at natural speed, without pausing or scrubbing. A text input option allowed them to add notes when no prede-

fined label fit. Although I initially envisioned cross-validation across multiple annotators to strengthen label reliability, operational realities made this impractical. I prioritized labeling by primary caregivers most familiar with residents, aligning with best practices emphasizing the importance of tacit knowledge in healthcare settings.

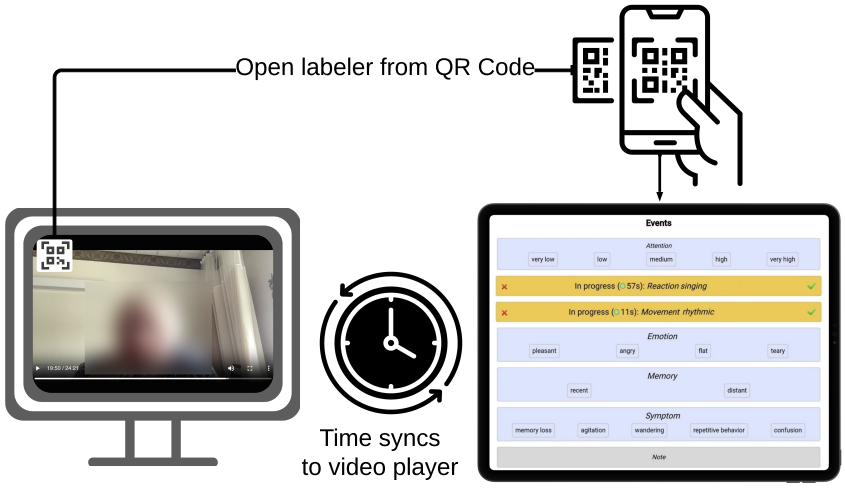


Figure 4.3: Caregivers use a two-device workflow to label sessions. Recorded videos are accessed on the Memcara secure cloud. Caregivers launch the labeling app by scanning a QR code displayed in the web platform, automatically syncing the labeling interface to the corresponding video session. Annotations are made in real time on a secondary device (e.g., tablet or phone), with engagement factors presented as parallel, toggleable tracks aligned to the video timeline.

Over a three-month period, staff annotated 203 video sessions, forming the supervision dataset. Throughout this process, model development proceeded in parallel with data collection. New labels were incorporated continuously to retrain and refine the multimodal engagement detection pipeline. From the outset, I envisioned the annotation process as a temporary bootstrapping phase rather than an ongoing burden. Once the AI system's engagement predictions achieved near parity with caregiver ratings, further labeling was discontinued.

This caregiver-annotated dataset formed the foundation for developing the automated engagement detection described in the following section. There, I describe the features

of the final CareInsights system: the solution for delivering more responsive, personalized dementia care within the constraints of a long-term care setting.

4.4 System Design

I developed CareInsights in close collaboration with caregivers to address the information breakdowns surfaced during the fieldwork described in Section 4.3. CareInsights is a modular infrastructure designed to embed sensemaking directly into long-term dementia care workflows. Rather than introducing new interfaces, CareInsights integrates with existing platforms (e.g., Memcara), supporting real-time capture, context retrieval, and narrative synthesis without adding burden to staff.

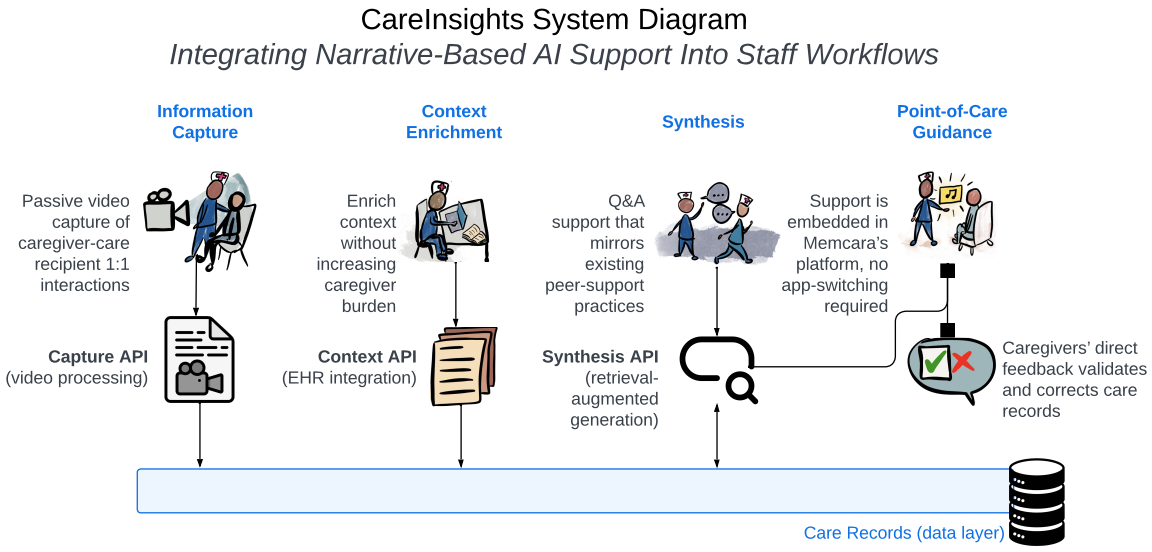


Figure 4.4: CareInsights system architecture, aligned with stages of the sensemaking loop. Each API supports a key aspect of caregiving information work: capturing meaningful observations (Capture API), enriching context from fragmented records (Context API), synthesizing insights into actionable narratives (Synthesis API), and delivering real-time, person-centered point-of-care guidance.

The system consists of three core APIs—**Capture**, **Context**, and **Synthesis**—which align with the sensemaking stages surfaced in the ethnographic research:

- **Capture API** identifies meaningful behavioral moments (e.g., affective shifts, redirection, engagement) from scheduled, tablet-based care sessions. This approach prioritizes emotional salience while respecting privacy by avoiding continuous video.
- **Context API** organizes fragmented medical data (e.g., medications, diagnoses, and incident reports) into coherent resident profiles, enabling rapid interpretation without manual cross-referencing.
- **Synthesis API** generates concise, natural-language answers to caregiver queries (e.g., "What helps calm this resident?"), combining structured records, past notes, and behavioral summaries into interpretable, actionable insight.

Insights are delivered contextually, whether during shift prep or in-the-moment escalation, thereby ensuring information appears where and when it is needed. This workflow-aware delivery supports decision-making without disrupting routines. Figure 4.4 illustrates how CareInsights maps to the stages of sensemaking observed in the field.

4.5 Implementation

This section describes the implementation of each CareInsights API and the approach to point-of-care guidance.

4.5.1 Capture API: Computer-Vision Pipeline for Video Summarization

Many meaningful insights in dementia care (e.g., how a resident feels, responds, or remembers) emerge during everyday interactions but often go undocumented. These moments are visible in video footage: a calming response to a familiar song, a memory sparked by an image, or distress triggered by a misunderstood cue. In the fieldwork, staff recalled music

soothing confusion, photos prompting stories, and container changes instigating food refusal due to past trauma. While such footage is rich with care-relevant signals, its volume makes it hard to review, leading to data overload. Without tools to surface and preserve key moments, insights are easily lost across shifts and teams.

The Capture API addresses this gap by embedding passive video-based observation into routines that already exist. Specifically, it builds on the Memcara tablet platform, which staff regularly use to deliver memory-oriented engagement sessions. Because tablets are naturally positioned to face residents during these sessions, they provide a consistent, stable framing of resident–caregiver interaction that is focused, unobtrusive, and repeatable across time and settings. This minimizes the technical and ethical barriers of prior work that relied on wall-mounted or ambient cameras, which struggled with occlusion, inconsistent angles, and privacy concerns. The system does not introduce new devices or behavioral changes, enabling high-fidelity naturalistic data capture.

4.5.1.1 Capture API Pipeline

During each session, synchronized video and audio streams are recorded automatically and processed through a multimodal pipeline. Visual features (e.g., facial emotion, drowsiness signals) are extracted using DeepFace and MediaPipe; transcripts are generated via Whisper and semantically enriched using GPT-4o. A stacked ensemble model (XGBoost) integrates these features to predict per-second engagement across five dimensions grounded in caregiver practice: attention, emotional expression, memory recall, reactions to therapy, and behavioral symptoms. Change-point detection is applied to these engagement signals to identify inflection points. Rule-based heuristics label each segment (e.g., a spike in memory recall or drop in behavioral symptoms), which are passed through a GPT-4o synthesis step. The result is a brief, natural-language summary that mirrors how caregivers document and communicate.

Figure 4.5 outlines the technical pipeline; Figure 4.6 shows example outputs. These results were used internally to support model development. While the capture layer focuses on structured behavioral observations, meaningful interpretation requires context—resident identity, baseline behavior, prompts used, and situational dynamics. To mirror this, the system continuously layers contextual information alongside engagement signals, supporting both real-time interpretation and downstream synthesis. Implementation details are provided in Appendix B.2.

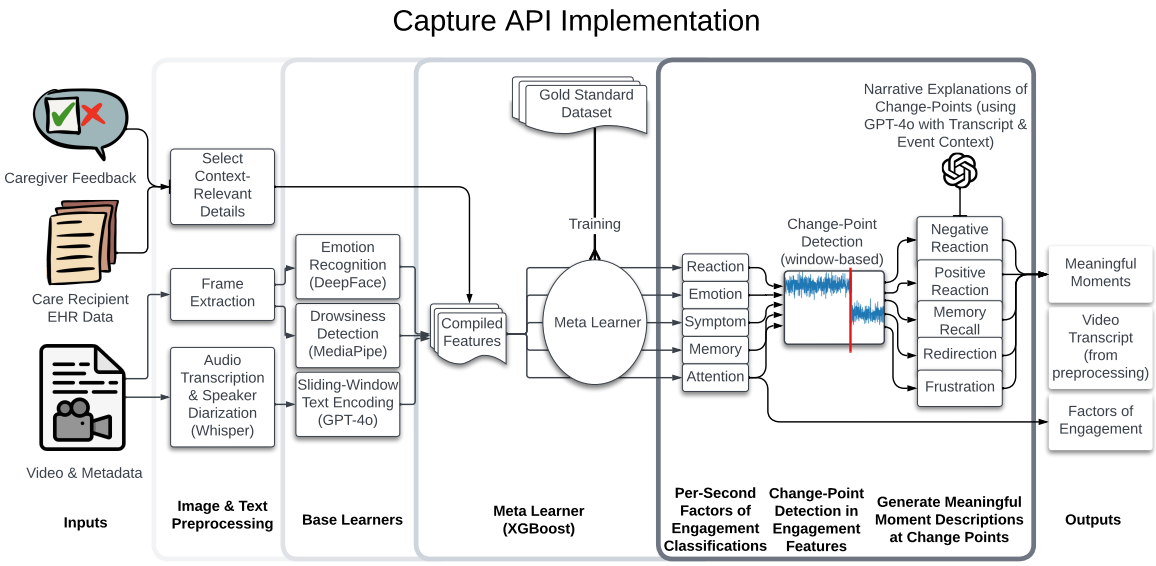
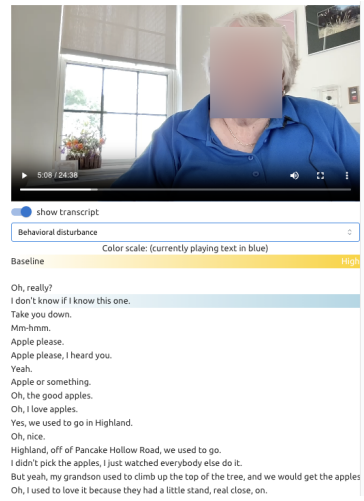


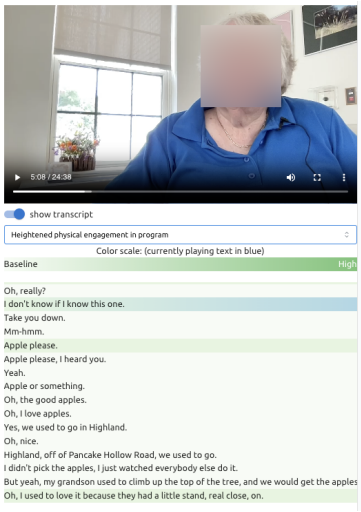
Figure 4.5: Given a video recording of a therapy session and relevant EHR metadata, the pipeline that supports the Capture API extracts visual and textual features using pretrained base models (DeepFace, MediaPipe, Whisper, GPT-4o). These are passed to a stacked ensemble learner (XGBoost), which generates per-second predictions across five engagement factors: attention, emotion, memory recall, reactions to therapy, and behavioral symptoms. Change point detection is applied to these signals to identify potential Meaningful Moments, labeled using rule-based heuristics and summarized using templated GPT-4o prompts. The resulting output includes time-aligned engagement trajectories, classified moment types, and narrative summaries suitable for documentation or review.

4.5.2 Visualizing Factors of Engagement

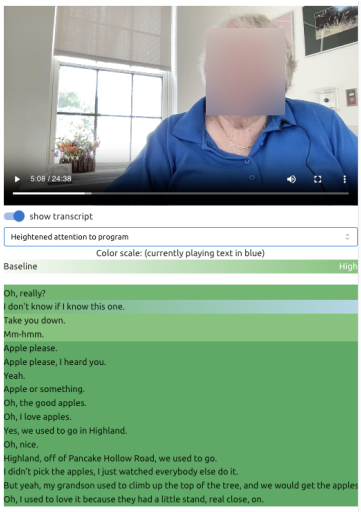
To evaluate and validate the Capture API's engagement predictions, my co-authors and I developed tooling that visualizes real caregiver–resident interactions from the deployed



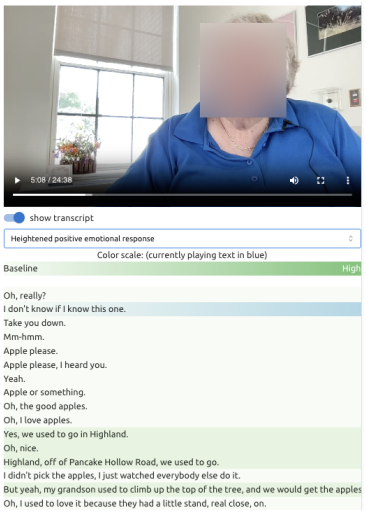
(a) Symptom: Behavioral Disturbance



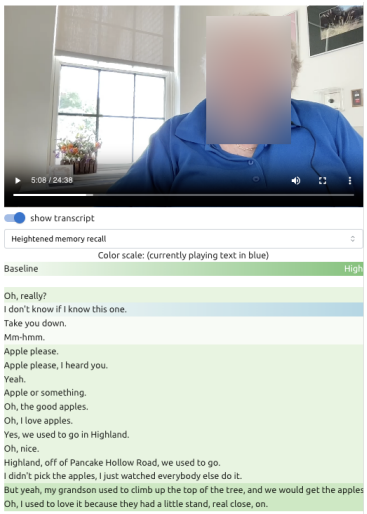
(b) Reaction: Heightened Physical Engagement



(c) Attention: Heightened Attention



(d) Emotion: Positive Emotional Response



(e) Memory Recall: Heightened Memory Recall

<Redacted Care Recipient Name> responds with visible positive emotion to an image of an apple orchard shown via Memcara's platform. She recounts a memory of visiting *<Redacted Location>* with her grandson *<Redacted Name>* to pick apples and enjoy fresh cider donuts.

Figure 4.6: Sample output from the Capture API. Five screenshots from the internal interface displaying a single moment in a caregiver–resident interaction, each showing a different engagement dimension selected by the user. For the same video segment, the system overlays predicted engagement levels and presents an automatically generated Meaningful Moment summary.

system. Each interface view overlays the session transcript with per-second predictions for a selected engagement dimension. For clarity during review, these predictions are averaged over the duration of each transcript block, aligning engagement signals with corresponding dialogue. In the example shown in Figure 4.6, a caregiver presents an image of apple trees via the Memcara platform, prompting the resident to recall her grandson climbing apple trees and exhibit visible positive emotion. This interaction was automatically detected and flagged as a Meaningful Moment, with the corresponding summary displayed in the lower right panel. This visualization tool was useful during development and testing for confirming that the engagement dimensions reflected caregiver-relevant interpretations of behavior. The following subsection describes how this context is structured to enable more person-centered insights.

4.5.3 Context API: Aggregating and Structuring Health Record Data

Understanding what occurs during care interactions requires up-to-date context—clinical, behavioral, and historical. The **Context API** maintains this contextual layer by transforming structured health records into narrative summaries that inform downstream interpretation and synthesis.

4.5.3.1 Context API Pipeline

The system pulls data from the Memcara platform, which serves as both a primary repository for EHR data and a secure bridge to other record systems. When relevant updates occur (e.g., changes in medications, behavior logs, or cognitive status), the Context API is triggered. Updates are parsed into structured fields grouped into domains like diagnoses, interventions, and care history and stored in a HIPAA-compliant back-end infrastructure shared with Memcara. From these fields, the system generates a short-form summary for lightweight synthesis and a long-form narrative for longitudinal reference, both timestamped and stored as context snapshots. These snapshots act as reference points during

interpretation, allowing the system to align synthesis with the care recipient's documented state at the time of the observed event.

These summaries are directly passed to downstream GPT prompts for both the Capture and the Synthesis APIs. By grounding each interaction in the most relevant context snapshot, the system ensures that outputs remain personalized, accurate, and temporally aligned.

4.5.4 Synthesis API: Context-Aware, Narrative-Based Clinical Guidance

To support sensemaking in real time, I introduce the **Synthesis API**, which generates concise, narrative explanations in response to structured caregiver queries.

Given a prompt such as "What should I know before seeing <*unfamiliar resident name*>?", "Has anything changed since yesterday's shift?", or "What's the best way to redirect <*care recipient name*> when he shouts?", the API retrieves relevant Meaningful Moments and other records from the data layer. It then synthesizes these snippets into a concise narrative that connects current behavior to past patterns, identifies likely causes, and offers practical guidance. Each explanation references the underlying evidence with citations, allowing caregivers to quickly verify details or explore supporting materials such as video recordings or EHR notes.

The narrative style was designed to mirror how caregivers communicate with one another in practice. During the fieldwork, I observed caregivers offering peer support through short, story-based tips that conveyed both practical advice and contextual understanding. The Synthesis API models this style, generating responses that balance brief narrative accounts with actionable suggestions. The specific level of detail, tone, and focus are determined by the calling context and passed in by the user, allowing the API to adapt its storytelling to the care recipient and situation at hand.

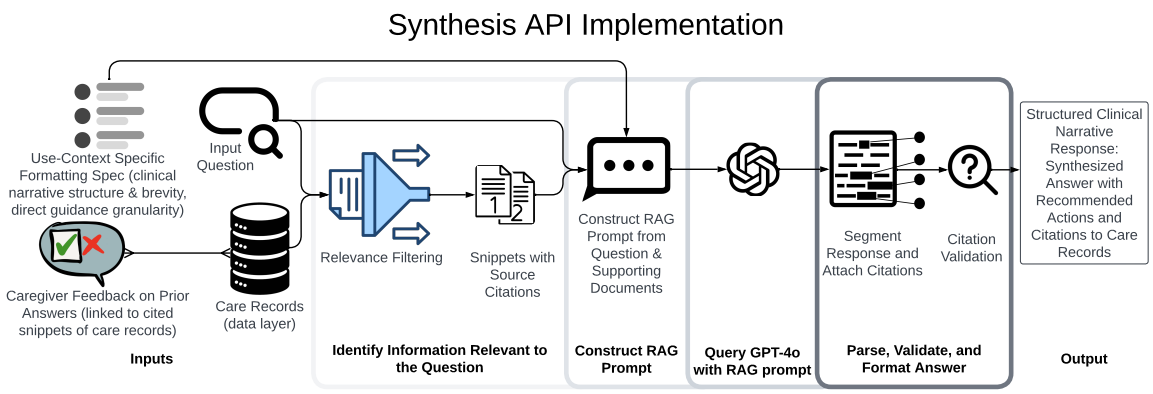


Figure 4.7: Overview of the Synthesis API pipeline. A natural language query is used to retrieve relevant documents, construct a context-aware prompt, and generate an explanatory, narrative-form response with supporting citations.

4.5.4.1 Synthesis API Pipeline

The Synthesis API uses a retrieval-augmented generation (RAG) pipeline—first retrieving relevant data, then conditioning the model's output on that evidence rather than relying solely on its internal parameters. RAG has been shown to improve personalization, factual grounding, and transparency by linking generated responses to specific, verifiable data sources [Lewis et al., 2020]. The Synthesis API draws from a shared data layer populated by the Capture and Context APIs, filters for relevant content using metadata and past feedback, and assembles a structured GPT-based prompt to generate the final response. The format of the output is exposed as an API parameter, allowing downstream tools (e.g., Memcara's platform) to request responses in styles suitable for handoff notes, just-in-time guidance, or formal documentation based on what is known about the use-case.

Each generated explanation includes citations linked to the underlying source records. Feedback on a response, such as corrections, affirmations, or additions, is automatically associated with the cited data, allowing users to vet system outputs and influence future retrieval behavior without needing to directly edit the data layer. This design enables lightweight, in-flow supervision of system behavior, while preserving trust and traceability. Figure 4.7 summarizes the pipeline; implementation details are in Appendix B.3.

4.5.5 Point-of-Care Guidance

Synthesized insights must ultimately reach caregivers in a form that meets practical requirements—i.e., timely, interpretable, and aligned with how care is actually delivered. This subsection describes how outputs from the Synthesis API are surfaced within existing workflows to support real-time understanding and response.

4.5.5.1 Designing for Generalizability and Interface Flexibility

I designed CareInsights as a general-purpose back end, intended to support a range of downstream interfaces rather than dictate a specific front end. Rather than constraining system outputs to a fixed format, I adopted a general templating strategy that accommodates varied display conventions and interaction models.

This modularity allows CareInsights to serve as infrastructure across diverse care environments. While this chapter presents a single interface view implementation, the system is designed to be adaptable, supporting different workflows, clinical tools, and display needs. The focus in this work is on demonstrating the value of structured synthesis and deep integration, rather than prescribing how that information should appear.

4.5.5.2 Integration with the Memcara Platform

Care work in long-term facilities is fast-paced, emotionally taxing, and shaped by constantly shifting priorities. In such contexts, I observed that a single format of feedback could not serve all roles or moments, and caregivers were generally reluctant to engage with additional standalone apps.

For the implementation, my co-authors and I integrated all three CareInsights APIs into the Memcara platform. This co-location allowed for seamless capture of data, sharing of information, and deployment. To keep the interface approachable and easy to adopt, a minimal interaction model was implemented: a simple question-and-answer chat that allows caregivers to submit natural language queries and receive structured, cited responses.

This format mirrors familiar messaging applications, minimizing the learning curve and making it suitable for use alongside existing documentation tools. While this integration demonstrates the back end's capability to support real-time decision-making, the system is not limited to this form. CareInsights is built to accommodate a wide range of delivery mechanisms, and future interface work may explore alternate visualizations, proactive notifications, or embedded documentation support.

Rather than privileging a single summary format, the Synthesis API treats narrative structure as an input parameter, enabling output to better align with the user's cognitive state and task demands. In the deployment, Memcara's platform inferred contextual state (e.g., leading a session, preparing for a shift, documenting a family meeting) from lightweight app metadata such as timestamp, activity state, and selected resident. During active sessions, the system prioritized concise, bullet-style outputs for quick comprehension, while other contexts surfaced longer vignettes with more detail and narrative richness. This adaptive framing supported situated sensemaking and actionable care insights by matching narrative format to cognitive context and task demands.

4.6 Deployment Study and Evaluation

My co-authors and I conducted a multi-method evaluation of CareInsights to assess its feasibility, usefulness, and fit for long-term dementia care. Methods included system analytics, validated instruments, surveys, structured caregiver assessments, and stakeholder interviews. Framed around the sensemaking process (capture, context, synthesis, and care actions), the evaluation examined real-world workflow integration and compared CareInsights to baseline tools like EHRs, paper notes, and verbal handoffs. The team interviewed 25 caregivers, family members, and healthcare professionals to understand system use and perception. Thematic analysis revealed key themes: improved multi-stakeholder commu-

nication, trust in AI-generated suggestions, and the value of integrating fragmented data for collaborative decision-making.

The evaluation was structured around each stage of the sensemaking loop. Table 4.2 outlines the metrics, methods, and justifications for each.

Table 4.2: Evaluation metrics by sensemaking stage with justification.

Sensemaking Stage	**Metric**	**Method**	**Justification**
Information Capture	Engagement detection accuracy	Leave-one-out cross-validation (video annotation dataset)	Accurate detection of behavioral cues is foundational to generating Meaningful Moments, enabling passive data collection without additional caregiver input. Validates technical soundness of the Capture API.
Context Enrichment	Utility, accuracy, completeness of video summaries (sequences of Meaningful Moments)	450 structured caregiver evaluations of generated summaries	Determines if automated summaries reflect clinical context and tacit knowledge needed for care continuity, documentation, and family communication.
Synthesis	Usefulness, accuracy, completeness of responses to Synthesis API queries	1,405 caregiver evaluations of AI-generated responses to care queries	Measures the ability of the system to surface relevant, timely information that supports planning and decision-making in dynamic care contexts.
Point-of-Care Guidance	Workload burden, engagement, implementation fit	NASA-TLX, UES-SF, AIM, IAM, FIM, person-centeredness, NPS	Evaluates if the system integrates into routines without increasing workload, is acceptable, aligns with person-centered values, and is feasible across facility constraints.

4.6.1 Survey Design and Administration

Matched pre- and post-deployment surveys were administered to caregivers across four of the seven originally planned partner facilities.[7] The baseline survey captured experiences with traditional tools and measured cognitive load, engagement, and the degree to which existing systems supported person-centered care. The post-deployment survey introduced additional instruments to evaluate the experience of using CareInsights. Post-deployment surveys added evaluations of CareInsights using validated instruments: NASA-TLX (The NASA Task Load Index) [Hart and Staveland, 1988], UES-SF (User Engagement Scale-Short Form) [O'Brien et al., 2018], AIM (Acceptability of Intervention Measure) [Weiner et al., 2017], IAM (Intervention Appropriateness Measure) [Weiner et al., 2017], FIM (Feasibility of Intervention Measure) [Weiner et al., 2017], and additional items on person-centeredness.[8] To gauge long-term scalability, the survey also included the Net Promoter Score (NPS), a single-item metric assessing willingness to recommend a product or service [Alismail et al., 2020].

4.7 Findings

The findings are presented below, organized by the corresponding stage of the sensemaking process.

4.7.1 Capture API: Performance Evaluation

My co-authors and I evaluated the performance of the Capture API using 203 manually annotated videos with leave-one-out cross-validation. Table 4.3 presents classification metrics

[7] Due to unanticipated administrative departures and funding cuts, several facilities faced substantial operational challenges. To avoid placing additional burden on staff at these sites, my co-authors and I continued supporting the CareInsights deployment while scaling back other research-related activities.

[8] For more information on these additional items, see Appendix B.5.

for five engagement categories: memory detection, attention intensity, emotion intensity, reaction intensity, and symptom intensity.

Overall, performance was broadly consistent with prior dementia-related engagement recognition benchmarks (see Section 4.3.3). These results suggest that the system's per-second predictions align well with caregiver-annotated reference labels. Given that these labels inform the construction of care-relevant narratives and Meaningful Moments (see Appendix B.1), and that the co-designed segmentation rules effectively map such predictions to usable highlights, these findings support confidence in the end-to-end pipeline under real-world deployment conditions.

Detection was strongest for engagement factors that were accompanied by clear verbal or behavioral signals. High performance on Symptom and Memory detection appeared to be supported in part by caregiver verbalizations during sessions. In many instances, symptoms were explicitly named, and distinctive keywords were captured in transcripts. Such narration inadvertently provides high-quality supervision signals for detection. Verbal cues, combined with expressive behaviors (e.g., smiling or singing—see Figure 4.6) further supported classification.

Table 4.3: Automated engagement detection performance metrics.

Factor	**Acc.**	**Prec.**	**Recall**	**F1**
Memory	0.93	0.90	0.93	0.91
Attention	0.61	0.70	0.60	0.60
Emotion	0.89	0.80	0.89	0.83
Reaction	0.82	0.75	0.82	0.76
Symptom	0.99	0.98	0.99	0.99

Emotion and Reaction were also reliably detected, often aided by caregiver commentary that clarified subtle responses. Attention intensity was the hardest to classify, as it often lacked clear cues.[9]

To interpret model feature importance across impairment severity, I analyzed audio-derived vs. visual-derived feature combined contributions as a function of MoCA across the care recipients with MoCA scores on file, using per-participant mean absolute SHAP magnitudes (not normalized to sum to 1). Visual features contributed most overall (mean 0.45–0.64). Both audio ($r = -0.78$, $p < .01$) and visual ($r = -0.45$, $p = .06$) importance decreased with higher MoCA, indicating greater reliance on perceptual cues at lower MoCA (greater impairment). This also suggests that the model leaned more on verbalizations at lower MoCA scores[10] (also increasingly from caregivers, considering that often PLWD with very advanced dementia are non-verbal), with visual cues remaining informative but less decisive. This pattern aligns with field observations and prior work on narrative scaffolding and reminiscence practices, where caregivers verbalize what residents appear to recall or feel, to sustain interaction and document care [Villar et al., 2019, Macleod et al., 2021, Steinert et al., 2020]. These results suggest that in cases of more advanced dementia, model performance may degrade in settings without caregiver narration. However, Steinert et al. [2020], whose work I use as a comparative benchmark, question whether striving for a strict separation between care recipient behavior and caregiver input is even appropriate. They interpret caregiver artifacts as natural and meaningful contributions that improve model performance, reinforcing the view that engagement in dementia care is fundamentally relational and shaped by contextual dynamics:

> *The notion of designing completely unassisted technical activation systems for people with (advanced) dementia may be misguided. Instead, the presence,*

[9] Field data confirmed caregivers also struggled to classify attention, especially with advanced dementia, unfamiliar residents, or in noisy, visually obstructed environments.

[10] The published version of this analysis [Moore et al., 2025a] reads "higher MoCA scores" here. That is a typographical error: the reported correlation ($r = -0.78$) indicates that audio importance rises as MoCA falls. The direction is corrected in this chapter.

contribution, and interest of tandem partners is not a methodological artefact or source of "noise" but rather an integral component of successful activation systems for people with dementia. Here, we showed how such a tandem-approach may yield ***meaningful moments of engagement***[11] *that can be successfully recognized via machine-learning techniques despite all challenges.* [Steinert et al., 2020]

4.7.2 Context API: Caregiver Evaluation of Generated Video Summaries

To assess how well the system surfaced clinically meaningful content, my co-authors and I analyzed caregiver responses to 450 structured evaluations of AI-generated Meaningful Moment summaries. These narrative form descriptions captured shifts in emotional expression, memory salience, or behavioral engagement that may warrant clinical or family attention. Each evaluation asked caregivers to rate the summaries along three dimensions: utility, accuracy, and completeness. These criteria were selected to capture the practical utility of the summary in care contexts (utility), its faithfulness to observed behavior (accuracy), and its coverage of relevant information (completeness). The choice of these dimensions was informed by formative interviews in which caregivers emphasized the need for clear, concise, and contextually grounded representations of resident activity.

[11] Boldface added to draw attention to Steinert et al.'s alignment with the terminology used in this thesis.

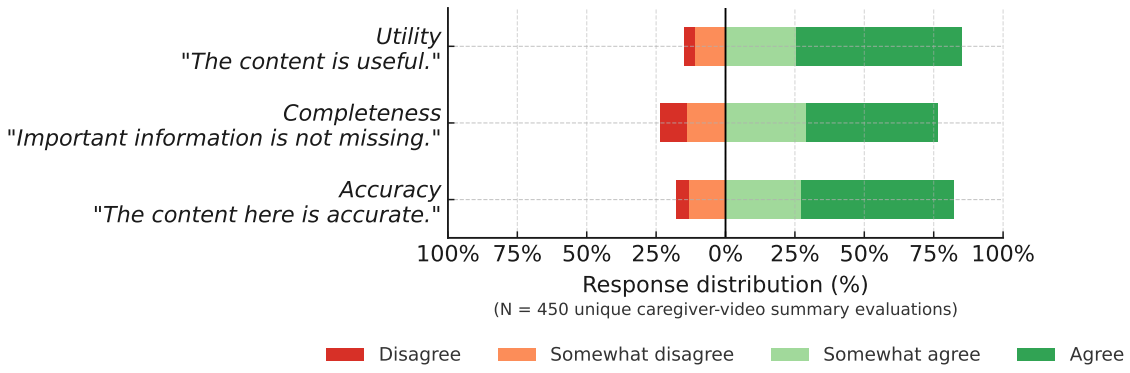


Figure 4.8: Caregiver evaluation of generated Meaningful Moment across three dimensions: accuracy, completeness, and utility.

Overall, caregivers tended to rate the summaries as useful, accurate, and complete, indicating that the system captured what they themselves considered salient and did not omit critical content. These results suggest that the moment generation pipeline produced summaries that were largely perceived as interpretable and actionable. When the team analyzed the subset of low-rated summaries, it was found that many came from early recordings affected by a platform-level audio-feedback issue in how the platform was recording video.[12]

To support the rating task, the team extended Memcara's platform with a timeline view—a feature that quickly proved valuable in its own right. I initially intended to prompt caregivers for real-time feedback during sessions, but this approach proved disruptive: it increased cognitive load and detracted from the tool's value as a support resource. Moreover, the person reviewing a video or receiving a query response was not always the staff member most familiar with the resident. Structured, post hoc evaluations were therefore redirected to staff with the most direct knowledge of the care recipient. These evaluators used the timeline view to access session recordings, transcripts, and a chronological list of Meaningful Moment summaries.

[12]Background music—especially songs with lyrics—were captured and transcribed, causing the system to misclassify non-resident audio as speech events. This occasionally produced false symptom predictions, such as attributing increased repetition of utterances—a clinical warning sign—to a resident when the "repetition" was in fact part of a song. The team identified and corrected the issue during the first month of gold-standard dataset labeling (see Figure 4.1). Thereafter, this error pattern did not recur, and caregiver evaluations stabilized.

Caregivers emphasized the utility of this format. Each Meaningful Moment appeared as a clickable item in the timeline, enabling direct navigation to the corresponding segment in the transcript or video. This tight coupling between summary and source supported rapid verification and validation, allowing staff to scan, confirm, and revisit specific events without scrubbing through lengthy media.

Recognizing its value, the team integrated the timeline into Memcara's production interface for reviewing recorded sessions. As F4-CG2 noted: "It's useful to have quick access to the information that matters without having to read through 10 pages of notes." Caregivers reported that the timeline was particularly helpful for preparing for family meetings and that seamless drill-down (e.g., from citations in feedback or from the session timeline) increased trust in the system-generated content. The linked summaries also had educational value for less-experienced staff, who appreciated being able to observe how more experienced colleagues delivered care without watching entire sessions. In these cases, Meaningful Moments functioned as narrative compression: providing just enough context to facilitate understanding without overwhelming the viewer.

4.7.3 Synthesis API: Caregiver Evaluations of Generated Responses

Caregivers evaluated 1,405 AI-generated responses produced by the Synthesis API, rating each response on utility, accuracy, and completeness. Figure 4.9 summarizes the results. Caregivers largely found the responses useful, especially for orienting to unfamiliar residents or summarizing behavior across time. Lower ratings were due not only to the complexity of summarizing care interactions, but also to technical limitations, such as the issue of misinterpreting background music as resident speech, which occasionally propagated into the synthesized responses.

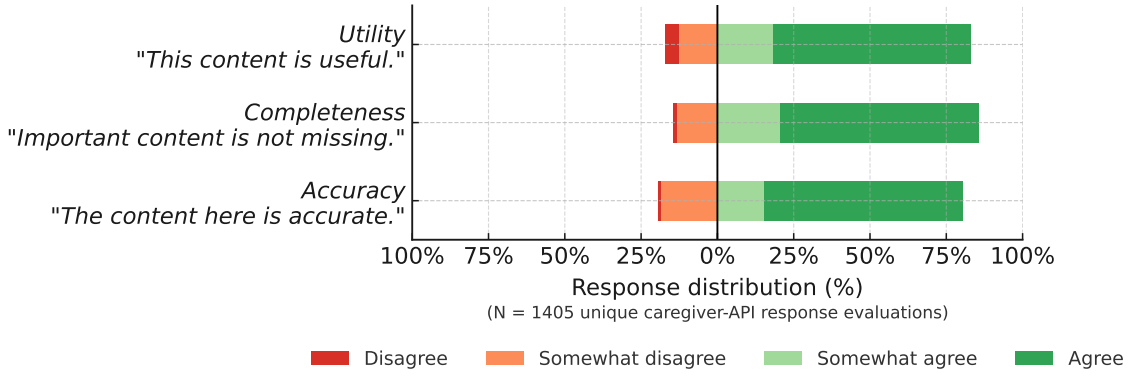


Figure 4.9: Caregiver evaluation of Synthesis API's generated responses (i.e., responses to user queries) across three dimensions: accuracy, completeness, and utility.

This highlights the need for faster detection and resolution of upstream issues, including more responsive filtering and improved video capture reliability.

Caregivers appreciated how the format of feedback matched their use-case. For instance, F3-CG4 described the benefit of bulleted responses while in the midst of 1:1 caregiving: *"I think the bullet points are probably enough most of the time, but especially when it comes to planning, the details are good to have there."*

Beyond pre-session and in-session support, caregivers used CareInsights to orient to unfamiliar residents, summarize behavioral patterns over time, and prepare for family meetings. The queries they posed to CareInsights reflected the same information needs and phrasing observed in caregiver-to-caregiver handoff conversations.

4.7.4 Point-of-Care Guidance Evaluation Measures

CareInsights was designed to surface actionable insights without adding caregiver burden. To assess its practical fit, my co-authors and I administered validated and adapted surveys across four long-term care facilities ($N = 19$ for the received survey responses). Constructs included workload, engagement, communication, person-centered care, and implementation feasibility. Unless stated otherwise, mean scores with standard deviations are reported.

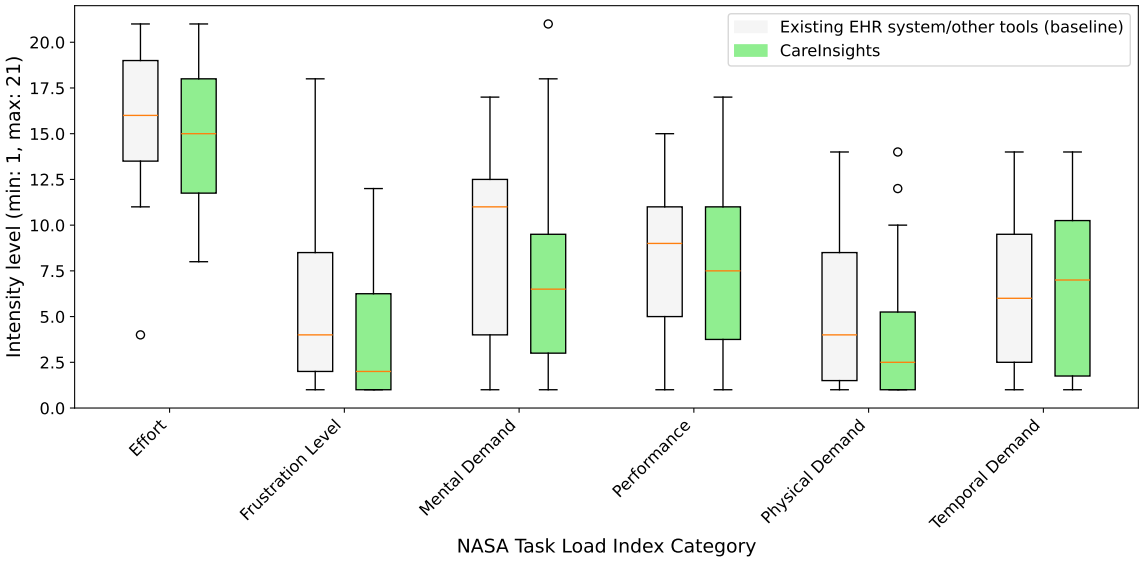


Figure 4.10: Pre/post study comparison of NASA Task Load Index scores across six workload categories. Lower scores reflect reduced frustration, cognitive burden, and time pressure. None of the differences in the subscales reached statistical significance (all $p > .05$).

Workload & Engagement: Median NASA-TLX workload scores (Figure 4.10) showed directional improvement in 5 of 6 subscales, but none reached significance (Wilcoxon signed-rank tests, all $p \geq .05$, Holm-adjusted $p \geq .49$). Given the limited power of the sample, these trends are interpreted as suggestive but not confirmatory. By contrast, NPS improved from -20 to $+60$ post-intervention. Engagement scores (UES-SF) rose from 28 to 32 post-deployment, though the change did not reach statistical significance ($p \geq .05$). This is interpreted as a directional trend.

Person-centered Care: Person-centered care scores increased significantly from 23.1 ± 5.8 to 36.8 ± 7.2 (out of 49), indicating improved integration of resident history and greater attention to individualized care. Additional detail on these survey items and subscales is provided in Appendix B.5.

Implementation Feasibility: Caregivers rated the system highly across standard implementation measures: AIM (Acceptability) = 17/20, IAM (Appropriateness) = 17/20, and FIM (Feasibility) = 16/20 (all SDs $\approx$ 4–5), indicating strong alignment with daily care routines and the feasibility of sustained uptake.

Usage and Integration: Over 400 sessions were recorded.[13] Most sessions lasted 20–30 minutes. Sustained use was enabled by seamless integration with the Memcara platform, avoiding new accounts or data entry, which the team learned was essential for adoption. Caregivers emphasized that they would not have used the system otherwise.

Caregiver Reflections: Caregivers valued immediate access to resident-specific context. Their trust was earned from both accuracy and the ability to verify or edit outputs. As F3-CG5 noted, *"I trust it because I can easily edit [via feedback] if it's not correct—and for the most part it was spot on."*

Summary: CareInsights supported sensemaking by delivering person-centered care insights within existing workflows. Passive video capture preserved rich context without increasing burden, and integration with familiar tools enabled adoption. Caregivers reported appreciating transparent, editable narratives that aligned with their own decision-making routines.

4.7.5 Evaluation Synthesis

Findings in Sections 4.7.1–4.7.4 collectively address the guiding research question. Section 4.7.1 shows that per-second engagement predictions support identifying Meaningful Moments. Section 4.7.2 demonstrates that Meaningful Moments are rated as being accurate, complete, and useful. Section 4.7.3 shows that RAG-based responses help caregivers orient and redirect with traceable citations. Section 4.7.4 indicates workflow fit (AIM/IAM/FIM), non-significant TLX subscales, and an NPS shift from -20 to $+60$. Building on these, three cross-cutting implications are synthesized below.

[13] Of these, 203 videos were manually annotated to support model development. These labels were only used for training purposes. For evaluation and live system operation, the team relied solely on system-generated labels so that synthesis reflected real-world performance rather than idealized outputs.

4.7.5.1 Alignment with Existing Infrastructure and Workflows Encourages Adoption

Sustained use was enabled by integrating CareInsights into a familiar platform with preloaded resident data. This avoided redundant setup and aligned with infrastructural and staffing constraints which surfaced during earlier fieldwork. Embedding CareInsights within an existing platform allowed the team to avoid common barriers to adoption such as new logins, duplicated data entry, or training requirements. By passively capturing and synthesizing interaction data, CareInsights reduced the need for extra documentation while providing timely, relevant context. Caregivers reported value in using the system during shift handoffs, family updates, and care planning. Caregivers emphasized that they would not have used the system if it had required more setup or operated outside their established routines. This supports the idea that designing for cognitive alignment offers a promising alternative to task-focused AI. It requires close collaboration with end users, attention to their real-world decision-making practices, and a shift from optimizing abstract performance metrics to supporting situated, context-aware utility.

4.7.5.2 Narrative-based Documentation Enhances Continuity of Care

The findings highlight the value of narrative as both a representational and reasoning tool in dementia care. Caregivers frequently used stories to explain behavior, justify interventions, and convey nuance across shifts. By aligning CareInsights with this mode of communication, the system was able to generate summaries that felt natural, interpretable, and actionable. These narrative outputs also revealed potential for broader use. Participants envisioned moment summaries informing quality-of-life tracking and serving as communication artifacts in interdisciplinary planning, suggesting that narrative-centered systems could help bridge qualitative and quantitative aspects of care.

Caregivers often make decisions based on tacit knowledge and accumulated experience under complex, fast-paced conditions. The results show that CareInsights supported

this sensemaking process by surfacing concise, narrative representations of behavior that mapped closely to how staff interpret, share, and act on information. These summaries were used to validate decisions, onboard staff, and communicate with families—filling gaps left by traditional documentation systems. Their usefulness stemmed not just from content, but from integration. Caregivers valued the ability to quickly trace summaries back to transcripts or video, increasing trust in the system and grounding interpretation in context. This underscores the potential for narrative-based outputs to serve as effective intermediaries between raw data and clinical insight, especially when designed with transparency and traceability in mind.

4.7.5.3 Cognitive Alignment Supports Interpretability

Caregivers described CareInsights as intuitive and relevant to their day-to-day needs, commenting that it "thought" like a caregiver would. This alignment made the tool feel interpretable and encouraged consistent use, reinforcing that perceived utility is not just about system accuracy but about how well the tool fits into cognitive workflows. The next section turns to a broader reflection on the implications of these findings.

4.8 Discussion

The findings illustrate how introducing a system like CareInsights into long-term dementia care can support the way caregivers generate, share, and act upon information within complex care environments. At its heart, this work seeks to demonstrate how the attentive ethos of person-centered care can be applied at an infrastructural scale with the use of context-aware, narrative-form, real-time caregiver support. By integrating lightweight video capture and AI-driven narrative synthesis into routine music-therapy sessions, CareInsights foregrounds relational knowledge that compliance-oriented documentation traditionally neglects, such as emotional nuances, biographical context, and subtle shifts in resident

engagement. This approach highlights how alleviating administrative workload can also support and encourage person-centered care.

4.8.1 Complex Adaptive Systems Perspective

To fully appreciate these contributions, it is useful to interpret long-term care facilities through the lens of complex adaptive systems, which has been widely applied across ecology, organizational studies, and health informatics. In this theory, systems are viewed as collections of interacting agents—caregivers, residents, family members, administrators, regulations, and technologies—whose interactions produce unpredictable, emergent behaviors [Holland, 1992, Mitchell, 2009]. Caregiving in these settings is marked by continual negotiation among competing stakeholder priorities, shifting resource constraints, and evolving resident needs. Thus, dementia care epitomizes what scholars call a "wicked problem" characterized by its inherent complexity, conflicting goals, and lack of definitive solutions [Rittel and Webber, 1973, Hannigan and McAllister, 2019, Alford and Head, 2017].

This "wickedness" is illustrated by the unintended consequences of well-intended attempts to improve care through increased documentation. Efforts intended to streamline quality assurance often paradoxically burden caregivers, decreasing their available time and emotional resources for providing nuanced, person-centered support [Anand et al., 2018, Hermann et al., 2022]. Recognizing these dynamics highlights the value of identifying what Meadows calls "leverage points" within these complex systems: small, strategically positioned interventions can be capable of catalyzing significant positive shifts [Meadows, 1999, Abson et al., 2017].

This work suggests that systems like CareInsights are such a leverage point. By capturing and surfacing relational insights that traditional compliance-focused EHRs neglect, the infrastructure enables caregivers to more proactively anticipate resident needs and coordinate responses. I position these findings alongside prior literature on storytelling, indicating

that the *narrative* aspect of shared documentation can strengthen the collective sensemaking abilities of care teams [Bansler et al., 2016, Sultanum et al., 2018, Gui et al., 2023, 2024].

4.8.2 Design Implications for Narrative AI Infrastructure

Although CareInsights was developed for dementia care in resource-constrained long-term care settings, this study may be informative more broadly as it illustrates how context-aware, low-burden tools can align with existing workflows and strengthen relationship-centered practice.

Drawing on principles of person-centered care, I propose the term *narrative AI infrastructure* to describe key qualities of the CareInsights system: a workflow-integrated, context-aware, and narrative-based approach to AI system design. This framing captures how heterogeneous data can be transformed into structured narrative scaffolds that support users' sensemaking, coordination, and relational understanding, while also minimizing documentation burden and preserving the contextual and interpersonal richness that narratives uniquely convey. The following subsections translate these qualities into design suggestions for building similar systems across other care contexts.

4.8.2.1 Represent Meaning in Narrative Form

Narrative-based documentation has long been recognized for its value in enhancing clinical decision-making [Greenhalgh and Hurwitz, 1999, Charon, 2001], sustaining person-centered dementia care [McKeown et al., 2010, Kindell et al., 2014], and guiding context-appropriate interventions [Lai et al., 2018]. These benefits arise in part from the narrative format's capacity to support interpretive work. By preserving temporal structure, causal relationships, and emotional and contextual nuance—while omitting extraneous detail—narratives offer a representational structure well-suited to capturing *meaning* as understood

through the lens of an informed observer. These same properties can also make narratives an effective format for conveying caregiving guidance.

However, as the fieldwork revealed, the cognitive demands of frontline care often limit caregivers' capacity to sustain narrative-based documentation. The effort required to compose, organize, and share narrative accounts competes with the immediate pressures of caregiving. When moments that caregivers consider meaningful are not recorded, data that could have been useful in future caregiving instances is at risk of being lost.

To address this challenge, narrative AI infrastructure can, first and foremost, serve as a well-informed notetaker—capturing key details of care encounters without adding burden. I argue that such systems should treat narrative as a first-class representational unit within system architecture. As the CareInsights approach demonstrates, short, contextually grounded stories can function as foundational elements for capturing, persisting, and synthesizing care knowledge. In CareInsights, this principle is realized through Meaningful Moments—brief narratives that integrate factual observations with interpretive context, serving as actionable knowledge artifacts to support documentation, coordination, and sensemaking. Representing information through narrative makes reasoning both structured and human-readable, enabling a RAG-based system like the Synthesis API to retrieve and synthesize narratives into targeted care guidance.

4.8.2.2 Model Domain Experts' Sensemaking

Narrative AI infrastructure that aims to represent meaning in care should also model how that meaning is constructed by domain experts. For example, the findings show that experienced caregivers first notice subtle changes in a resident's engagement, interpret those shifts within the person's biographical and clinical context, and infer what they imply for care. CareInsights mirrors this sensemaking process by structuring inference as a sequence of interpretable stages rather than a single end-to-end prediction.

While this architecture introduces tradeoffs (see Section 4.8.3.3), it affords a key benefit: intermediate judgments (e.g., engagement factors and detected Meaningful Moments) remain visible and interpretable in ways that align with caregiver reasoning. In this case, caregivers saw CareInsights as reflecting their own sensemaking practices and appreciated the ability to inspect and correct outputs when needed. This transparency supported a stronger sense of agency and contributed to more positive attitudes toward the system overall.

To support cognitive alignment, I recommend structuring computational processes to surface interpretable reasoning steps, draw on constructs familiar to users, and maintain clear links to underlying evidence.

4.8.2.3 Design for Workflow Fit

Narrative AI infrastructure should fit naturally into existing caregiver workflows. Both prior work and field observations highlight how challenging it is to introduce new technical systems in long-term care settings. Caregivers work under high time pressure and often have limited tolerance for workflow disruptions. In response, CareInsights was designed to integrate into existing practices, EHR systems, and tools (i.e., Memcara's platform). In the fieldwork, I observed caregivers sharing stories about care recipients as a means of care guidance. CareInsights extends this storytelling practice at an infrastructural level by capturing and surfacing such narratives within everyday workflows. This approach represents my interpretation of workflow fit.

Understanding and aligning with caregivers' workflows required both relational and technical groundwork. Through fieldwork and collaboration with Memcara, my co-authors and I engaged regularly with care teams to develop rapport and better understand their routines. On the technical side, the team integrated with existing tools and records with Memcara's support. Based on these experiences, I encourage collaborating with industry partners and allowing time for fieldwork and technical integrations.

4.8.2.4 Design for Context-responsive Narrative Formats

As discussed earlier, narratives offer a flexible and intuitive way to communicate care knowledge. They can be stretched, condensed, or reorganized to support different types of sensemaking and communication. Designers of narrative AI infrastructure should leverage this flexibility to adapt care guidance provided by the system to users' cognitive and situational needs.

For example, the Synthesis API treats narrative format as a configurable parameter (see Section 4.5.5.2), allowing downstream interfaces to request narrative-based guidance tailored to the task at hand.

4.8.2.5 Set Ethical Boundaries on Data Collection and Use

In dementia care, residents may display behaviors—such as aggression or disinhibition—that do not reflect their premorbid identity. How these moments are documented and communicated (including to family members) can influence how individuals are perceived and supported. While these concerns are especially salient in this context, narrative AI systems more broadly must consider privacy, dignity, and clinical relevance when determining what to capture, what to forget, and how to interpret events.

I argue for framing narrative AI infrastructure as a situated co-storyteller: an intentional actor that advocates for and negotiates between the needs of individuals within a community. Especially in contexts like this one, where surveillance and automation risk undermining trust, it is important for designers to define clear boundaries around data collection. In this framing, the system is accountable for both what is sensed and how those experiences are represented.

In CareInsights, the team operationalized boundaries that caregivers and families requested through session-bounded sensing, limiting capture to pre-scheduled music therapy sessions. This created a well-defined and consented scope of use, helping staff and families feel that data collection was purposeful and respectful rather than intrusive. My co-authors

and I also collaborated closely with caregivers throughout the design and development process to ensure the system reflected their values, needs, and expectations. I believe this user-centered process can help align narrative AI infrastructure with community values and caregiving practice.

4.8.3 Limitations and Future Work

4.8.3.1 Longitudinal and Population-level Use

Beyond the session-focused care insights explored in this work, staff expressed interest in having dedicated longitudinal and population-level views—for example, tools to track or predict changes in resident engagement over time, or to compare the effectiveness of interventions across groups. Future work should investigate how narrative AI infrastructure might support these needs, including opportunities to improve activity programming at facilities, tailor and evaluate residents' care plans, and identify emerging trends across many residents.

4.8.3.2 Measuring Trust

The evaluation surfaced indicators of trust, including the perceived value of citations and transparent system reasoning. However, trust in narrative AI infrastructure should be more deeply explored. For a system like CareInsights, for instance, open questions include: When do caregivers choose to rely on narrative guidance? Where does trust break down, and why? And how can system design support calibrated, discretionary trust rather than blind acceptance or habitual dismissal? A deeper understanding of these dynamics may help guide safer, more ethical, and sustainable adoption.

4.8.3.3 Model Improvements

For this initial field deployment, a stage-based, interpretable architecture was adopted to emphasize traceability, operational predictability, and cognitive alignment with caregivers' sensemaking. While this was sufficient for the present use case, a natural next step is to incorporate multimodal foundation models for audio and video via modular adapters or temporal encoders. This could strengthen cross-modal alignment and generalizability across settings, while enabling newer backbones to be swapped in with minimal retraining. In the long run, using foundation models as end-to-end learners rather than only as feature extractors is likely preferable for maintainability and performance. However, this introduces interpretability tradeoffs that future work could also explore.

4.8.3.4 Establishing Baselines

Despite growing interest in AI-enabled documentation, work on person-centered dementia care still lacks the canonical benchmarks that have propelled progress in other areas of medical AI. As discussed in Section 4.3.3, existing benchmarks underrepresent the relational and narrative subtleties of this domain. An important area for future research is creating shared benchmarks for person-centered dementia care.

4.8.3.5 Involving Families

This study primarily focused on professional caregivers, leaving out other essential perspectives—particularly those of family members and clinicians—from both the design and evaluation of CareInsights. However, these individuals play a critical role in shaping how care is delivered. Dementia care literature has long emphasized the importance of sustained family engagement with facilities in fostering transparency, shared understanding, and resident well-being [Hayward et al., 2022]. Future work should seek to more actively include family members and other care collaborators as co-design partners, examining how narrative infrastructures might foster trust and care continuity. Although interface

design was not a central focus of this study, broadening participation should also inform how systems are presented—ensuring that outputs remain accessible, interpretable, and meaningful to all those involved in the care network.

4.8.3.6 Broader Next Steps

This chapter shows how narrative AI infrastructure like CareInsights can support person-centered dementia care. Future work can build on this foundation to explore its relevance to other domains where relational work, narrative understanding, and contextual reasoning are also important.

4.9 Conclusion

Dementia caregiving in resource-constrained long-term care facilities demands continuous, high-stakes sensemaking. Caregivers observe resident behavior, interpret it through prior experience and clinical context, and act under time pressure. In this context, documentation about care recipients could play a more active role in supporting person-centered care practices than it currently does. I designed CareInsights, a workflow-integrated, AI-enabled system to make documentation actionable by pairing passive sensing with adaptive, narrative-based feedback that supports person-centered dementia care. Across multiple long-term care facilities, my deployment showed that CareInsights fit existing workflows, supported sensemaking, and flexibly addressed the relational complexities of dementia care, with no significant change in measured workload. I conceive of CareInsights as a form of "narrative AI infrastructure" and outline next steps for advancing generalizable, narrative-based systems that support person-centered care at scale. Chapter 5 takes up one of those next steps, extending CareInsights to family members through Family In The Loop.

Chapter 5

Family In The Loop: Enabling Family Involvement in Dementia Care

This chapter builds on CareInsights (Chapter 4) and is adapted from [Moore et al., 2025b].

5.1 Introduction and Motivation

As the global population ages, the number of people living with dementia continues to rise, with estimates projecting an increase from 57 million in 2021 to 78 million by 2030 [World Health Organization, 2023, 2021]. Long-term care facilities, already stretched thin, are under increasing pressure to meet the complex and evolving needs of this population [Shih et al., 2014]. Unlike general geriatric care, dementia care involves navigating a wide range of neuropsychiatric symptoms—from cognitive decline and behavioral disturbances to progressive loss of function—that often lead to near-total dependence in the later stages of the disease [Snowden et al., 2003, Kratz, 2017, Kverno et al., 2009].

Person-centered care is widely recognized as the gold standard in dementia care. It views the resident as a person with an enduring sense of self, unique preferences and personality, and seeks to support this personhood [Manthorpe and Samsi, 2016]. Evidence-based therapies such as reminiscence, validation, and cognitive stimulation reflect this ap-

proach and offer pathways to more meaningful engagement. However, in practice, these therapies are difficult to sustain. Delivering them requires time, emotional energy, and deep contextual knowledge—resources that are often scarce on the ground. Understaffing, burnout, and high turnover leave caregivers overextended and reactive, rather than proactive and prepared to administer tailored care [Gandhi et al., 2021].

These constraints are further compounded by information breakdowns. As dementia progresses, PLWD lose the ability to explain what helps, what harms, or how they're feeling—placing even more pressure on staff to interpret subtle signs without consistent guidance. Meanwhile, gaps in documentation, siloed communication, and underutilized family knowledge disrupt continuity of care [Cadieux et al., 2013, Vernooij-Dassen et al., 2010]. The result is a persistent gap between what high-quality, person-centered care promises—and what the realities of care work allow.

Emerging digital interventions, such as mobile apps, web portals, and AI-driven assistants, offer promising avenues for meeting this need [Wang et al., 2024a, Cherian et al., 2024]. In particular, digital agents (software tools designed to perform tasks, gather information, and interact with users either fully or semi-autonomously) have emerged as an active area of research in health care contexts, including long-term care facilities [Tudor Car et al., 2020]. Such tools offer versatile solutions that help facilities manage complex workflows, support caregivers in resource-constrained facilities, and improve resident well-being [Sun and Zhou, 2023, Lindeman et al., 2020]. However, implementing such technologies for PLWD requires careful attention to unique challenges of this context and the condition itself, including issues of shared caregiving responsibilities as well as privacy, legal, and ethical considerations [Mujirishvili et al., 2023].

Another critical challenge relates to supporting family involvement in care practices. Although existing dementia care models and digital interventions integrate biological, psychological, and social factors and acknowledge the crucial role of family members in the care of PLWD, they often overlook the role of the care recipient's family system and rela-

tionships with caregiving staff as determinants of care outcomes [Podgorski et al., 2021]. Moreover, little research has examined how technology-driven information-sharing affects family–facility dynamics. To address this gap, this work adopts a more inclusive dementia care framework that emphasizes not only person-centered care but also a family-centered lens [Podgorski et al., 2021, Hayward et al., 2023], which considers the relationships within the social ecological networks that encompass care recipients, families, and caregivers. This chapter specifically explores a gap in the literature on the design of artificial intelligence tools that promote family engagement in person-centered dementia caregiving at long-term care facilities.

By providing scalable solutions for real-time information sharing on resident status, care plans, and facility decisions, such technologies could potentially transform both care workflows as well as families' ability to stay connected and informed [Zhai et al., 2023]. For example, a well-designed digital portal could give families timely, accurate insights into their loved one's care, thereby fostering trust, alleviating anxiety, and allowing them to feel actively involved in the decision-making process without imposing additional burdens on caregiving staff.

This chapter presents the needfinding, design, development, and evaluation of **"Family In The Loop"**, a web-based system that supports person-centered dementia care by bridging communication gaps between families and care staff. FITL supports family-staff collaboration by tailoring shared access to care information through four interlinked views that provide families with emotionally attuned visibility into the loved one's experience while enabling caregivers to retrieve essential information on demand, all without disrupting existing workflows. By supporting the capture, curation, and sharing of care-related information, FITL addresses the entangled emotional, clinical, and contextual needs of users. Through a multi-year, multi-site collaboration, my co-authors and I identify staff and family-specific needs, navigate institutional concerns around documentation and liability, and evaluate a deployed prototype. This work contributes new insights into how

sociotechnical systems can enhance visibility, trust, and collaboration in long-term dementia care.

Specifically, the efforts presented in this chapter were guided by the following research questions that reflect that process:

- RQ1: What are the experiences of families and professional caregivers of care recipients living with dementia in long-term care facilities, particularly with respect to challenges in communication and family involvement?
- RQ2: How might we design and build an interactive tool that addresses the needs of family members and caregivers and enhances relationship-building to foster a greater sense of engagement for families of PLWD?
- RQ3: How does the tool impact families' sense of involvement in care as well as other caregivers' practices?

5.2 Background and Related Work

This section reviews foundations for this case study, focusing on how person-centered care, care ethics, and sensemaking manifest in the specific context of family involvement in dementia care.

5.2.1 Person-Centered Dementia Care in Long-Term Care Facilities

Numerous medications exist to address the behavioral and psychological symptoms of dementia; however, these treatments often lead to unwanted side effects and have limited effectiveness [Dyer et al., 2018, Kolanowski et al., 2017]. Overmedication of psychotropic drugs in particular is common among PLWD at care facilities and is a point of concern for families and staff due to risks for adverse drug effects, drug interactions, and negative impacts on quality of life [Flamm, 2018]. In response, there is a growing shift in dementia

care towards non-pharmacological interventions that prioritize cognitive training, physical activity, and behavioral techniques [Brett et al., 2016, Hill et al., 2017]. These strategies aim to promote the abilities and potential of PLWD, avoiding a pathologized view of their condition. Consequently, the term "care recipients" is preferred over "patients" to reduce associations with pathology and deficit-based labeling.

As discussed in the previous chapter, person-centered care goes beyond treating symptoms and emphasizes understanding a PLWD's history, preferences, and values. To deliver such care, prior research highlights the importance of respecting a care recipient's preferences, providing education and emotional support, and involving family members in their care [Rathert et al., 2012]. By strengthening interpersonal connections and alleviating stress, this approach facilitates a more supportive caregiving environment. Although caregivers have the desire to foster greater communication with care recipient's families, this is challenging due to cumbersome documentation, staff shortages, high workload, and increased attrition rates [Hermann et al., 2022]. AI-driven digital health technology has the potential to provide valuable contributions towards mitigating the aforementioned problems by enhancing personalized communication and increasing support for caregivers [Wang et al., 2024a].

5.2.2 Family Involvement in Care

Even after a PLWD is admitted to a long-term facility, family caregiving continues [Rathert et al., 2012]. However, the type of care evolves: families transition from direct physical assistance to building relationships with care staff, offering valuable insights that could greatly enhance care, and participating in care planning to discuss needs and decisions regarding a loved one's condition [Gaugler and Mitchell, 2022, Bhat et al., 2023]. Yet, barriers to family-involvement remain, including strained staff–family relations [Gaugler and Mitchell, 2022, Puurveen et al., 2018]. Meaningful family participation is more likely when communication is regular and transparent, interactions are grounded in respect and

trust, and staff explicitly acknowledge family needs and concerns [Barken and Lowndes, 2018].

Families regard communication as a means to foster collaboration between themselves and the staff, aimed at enhancing the resident's well-being [Bauer and Nay, 2011]. However, significant barriers can still make family members hesitant to engage in a loved one's care. These barriers include insufficient support from care staff or management, strained staff–family relationships, reluctance to embrace change, and other administrative challenges [Haesler et al., 2007]. Family satisfaction has been shown to depend on how much family members feel they are actively involved by the facility in delivering personalized care to their loved one [Reid et al., 2007]. Another study that measured the relationship between family involvement and family perceptions of the quality of life of care recipients in long-term care facilities demonstrated that frequent communication between families and facility staff resulted in higher family perceptions of resident quality of life. Therefore, facilities should be encouraged to become more family-oriented through encouraging communication between staff and families [Podgorski et al., 2021].

Family involvement is essential for delivering the best care for individuals with dementia, especially when it comes to personalized care. This personalized approach is a key element of high-quality services provided by long-term care facilities [Reid et al., 2007]. Despite this recognized significance, there is a scarcity of interventions that address the communication and relationship-building challenges faced by caregivers and family members [Backhaus et al., 2020]. Additionally, few studies explore the potential of technological interventions to alleviate the resource limitations experienced by staff and to address these specific challenges. This work contributes to the literature by inquiring into the role of family–staff communication in long-term care, and it presents findings that underscore the importance of engagement between families and care providers for residents' well-being.

5.2.3 Family-Centered Digital Caregiving Tools

Despite rapid growth in healthcare technology to support caregiving [Cherian et al., 2024], their potential remains underexamined when it comes to enabling family involvement in dementia care within long-term facilities. In particular, there is a gap in the literature regarding what information families want about a loved one's health and status, how they would like to receive these insights, and how this information would affect their involvement in the person's care. Closing this gap is essential for designing family-centered tools that fit resource-constrained workflows rather than adding burden.

When not properly facilitated, overburdened staff can experience family interactions as an added burden, leading to avoidance, perfunctory updates, or strained relationships [Utley-Smith et al., 2009]. Evidence-based strategies can reliably foster constructive engagement, particularly relationship-centered practices that honor residents' uniqueness and share concrete experiences of care [Haesler et al., 2007]. Yet there remains no consensus on the content and structure of routine staff–family communication, underscoring the need for tools that are sensitive to both established strategies and individual preferences, especially in settings that are already constrained on resources including caregivers' emotional bandwidth.

A systematic review found that among the 29 studies on interventions to foster family inclusion in nursing homes for care recipients, only two aimed to create family-staff partnerships, and only eight focused on enabling family members to make better-informed decisions and/or participate more actively in future discussions with staff [Backhaus et al., 2020]. These studies were limited in their scope, mostly concentrating on family members with regard to their decision-making in end-of-life care rather than measuring whether the interventions led to a greater perception of family inclusion in care. Therefore, interventions should consider a broader, more holistic, and person-centered role for family members of people with dementia in long-term care facilities. In all the surveyed literature, reviews, and meta analyses, interventions mostly consisted of staff training programs, fa-

cilitated family conferencing, compassion-based care programs, and toolkits and booklets to aid conversations. Scarcely any interventions took aim at examining how digital tools and platforms could potentially enhance staff information sharing and provide updates on patient care to families.

One promising study investigated the complex and challenging dynamics of family involvement following a care recipient's admission into long-term care facilities through the promotion of communication opportunities via an individualized technological database. This database consisted of personal favorite music, photos from meaningful life events, and special souvenirs and other artifacts that were collected by the spouses of the care recipients. The personal material was found to encourage engagement and social interaction by assisting the spouses to find ways to communicate with their partners living with dementia, re-engage with shared past experiences, and reduce stress and feelings of disconnection during visits [Dassa, 2018].

This work seeks to understand and improve the state of communication not just among caregivers and recipients (i.e., communication between staff members, or between staff and PLWD) but also communication between staff and family members of the care recipient. As illustrated by prior interventions, effective information sharing cultivates trust between families and care facilities, alleviating family anxiety and strengthening relationships with the care team. Going further, the research presented in this chapter incorporates human-centered design to investigate how technology can minimize communication barriers, enhance family satisfaction with facility staff's caregiving, and promote more person-centered care by enabling new avenues of family involvement.

5.3 Method

In this section, I outline the human-centered design approach, the participants and care facilities with whom my co-authors and I collaborated, and the methods and tools used. All

protocols were reviewed and approved by the Institutional Review Board (IRB) at Dartmouth College.

5.3.1 Design Framework

To orient the research process, my co-authors and I took an iterative design approach ("Double Diamond") [Design Council, 2021] through the lens of Podgorski et al. [2021]'s biopsychosocial-ecological framework. This framework emphasizes the integration of biological, psychological, social, and ecological factors in dementia care, therefore incorporating a complex understanding of the coordinated efforts of families, professional caregivers, and residents themselves. This approach grounded us to better understand individuals with dementia in relation to their family dynamics and social networks, facilitating a holistic perspective on their personal and relational needs within the broader dementia care context.

In addition, we approached long-term care facilities as complex adaptive systems, or ecologies, as described in Section 4.8.1. This perspective helped us anticipate the inherent challenges involved in designing scalable, person-centered digital interventions. Specifically, we recognized that emergent behaviors among various agents—both human stakeholders (e.g., caregivers, residents, families, administrators) as well as digital agents—are dynamic, nonlinear, and constantly adapting based on interactions. Under such conditions, interventions that cannot adapt to changing contexts and variable patterns of user engagement tend to lose their effectiveness [Carroll et al., 2023, Chaffee and McNeill, 2007, Gilman, 2021, Pype et al., 2018]. To address these concerns, we took an expansive view of our user bases, engaging a sample of users that covered nearly the complete care ecosystem at all stages of our process.

We employed a four-phase Double Diamond design process (**Discovery**, **Definition**, **Development**, and **Delivery**) to investigate, design, and evaluate our intervention aimed at enhancing family involvement in dementia care at long-term facilities while supporting

staff caregiving practices. Different stages addressed our three research questions (RQ1, RQ2, RQ3—all outlined at the end of Section 5.1). Specifically, Discovery centered on understanding the experiences and challenges of families and caregivers (RQ1), Definition and Development addressed how we could design and implement interactive features to meet the identified needs (RQ2), and Delivery focused on real-world system deployment and evaluation of its impact on family engagement and staff practices (RQ3). Overall, our approach seeks to create a technology that enhances family engagement through transparent, accessible communication within dementia care facilities. The following sections detail the methods used in each phase, describing participant recruitment, data collection procedures, system development, and analytic approaches.

5.3.2 Clinical Collaborators, Care Settings, Participants, and Recruitment

5.3.2.1 Collaborations with Care Providers

To ensure the development of dementia care technology that is directly grounded in real-world experiences, needs, and domain expertise, my co-authors and I partnered with a team of memory care providers from Memcara, which specializes in person-centered care and offers in-person and web-based services to care facilities across North America.

5.3.2.2 Care Settings

To identify potential care sites, we leveraged Memcara's connections to long-term care facilities. We also relied on Memcara's determination of whether a facility was resource-constrained, based on its access to critical resources (staffing, funding, equipment, training), whether the facility was struggling to meet enrollment requests and family needs, and its ratio of residents to staff[1]. Ultimately, we partnered with seven care facilities located

[1]A "resource-constrained" long-term care facility is one in which the demand for staff time, expertise, and equipment routinely exceeds the available supply (e.g., due to high resident loads, high symptom acuity,

across the U.S. and Canada, which we refer to as F1–F7. As is common for long-term care facilities, none of our participating facilities specialized in dementia care, but each included residents who were older adults living with moderate to severe dementia.

5.3.2.3 Participants and Recruitment

Through collaboration with facility administrators at F1–F7, we recruited 40 professional caregivers (including nurses, certified nursing assistants, and therapeutic recreation specialists), 35 PLWD as care recipients, and one representative family member per care recipient for a total of 35 family members. Following our IRB-approved enrollment procedures for a study involving older adults and PLWD, we obtained written informed consent or proxy consent where applicable. Family members were contacted via facility-wide email announcements, posted flyers, and word of mouth; staff caregivers were recruited with the support of each site's administrative team.

5.3.3 Discovery (Addressing RQ1)

In the Discovery phase, we sought to answer RQ1 by exploring how families and caregivers experience day-to-day communication; what barriers they encounter when capturing, using, and sharing information about the PLWD; and what opportunities might exist for improving family involvement.

Field Notes I carried out a year-long observational study, conducting weekly site visits across all seven facility locations to observe and document care practices and interactions among staff, residents, and families in real-world contexts. The observational field notes captured issues such as staff workload, technology usage, and the pain points and emotional tone of impromptu family visits and quarterly family-staff care conferences.

or shortages of trained staff). This creates chronic capacity strain due to insufficient personnel or resources to meet residents' needs, leading to overworked staff and a limited ability to provide optimal care [Gandhi et al., 2021, Anesi and Kerlin, 2021].

Interviews In parallel, my co-authors and I conducted 16 semi-structured interviews, each lasting 60 to 90 minutes, with family members and staff caregivers. The semi-structured format allowed for flexibility, enabling participants to share in-depth personal experiences while still addressing key themes we sought to investigate. When possible, interviews were conducted in person at the facility, but some interviews were conducted over Zoom due to facility location or scheduling constraints. Questions explored how participants currently manage communication about a resident's health and routines, the types of information families most value, and any resource or regulatory constraints that shape staff–family collaboration. Interviews were recorded when permission was obtained to do so.

5.3.4 Definition (Addressing RQ1 and RQ2)

We digitally transcribed all interview audio recordings and consolidated our field notes from observations into a shared repository. Three coders each performed in vivo coding to capture participants' own language around perceived communication gaps, staff scheduling pressures, and family members' desire for real-time updates. Our synthesis primarily occurred during weekly research team meetings, which included representatives from Memcara. During these meetings, we reviewed and refined our coding categories based on recent observations and identified emergent themes, such as how short staffing and privacy regulations limited staff–family engagement. From the refined thematic clusters, we subsequently synthesized them into key insights about how information on PLWD is recorded, shared, and used to provide care and create a sense of involvement for family members. We compared our own findings to the established literature on family involvement in care, noting agreement, differences, and enrichment. To enhance clarity and coherence in this chapter, the presentation focuses on results related to information-sharing practices.

Building on these findings captured during the Discovery phase, our Definition phase aimed to translate identified challenges into concrete design priorities. This laid the ground-

work for RQ2 by conceptualizing the types of interactive features and data-sharing mechanisms that could address our identified themes. Finally, we derived specific functional requirements for our intervention.

5.3.5 Development (Addressing RQ2)

We next developed low- and high-fidelity prototypes to explore technological features aligned with our design priorities. We called our system "Family In The Loop" or "FITL". Evocative of the phrase "human in the loop" [Monarch, 2021], FITL's name is intended to capture the ideas of human-AI interaction as well as collaboration between facility staff and family members. Complementing the methods presented here, specific user feedback and general themes presented in Section 5.4 further clarify how the core challenges identified in RQ1 drove our design goals and iterative development process.

5.3.5.1 Initial Design Goals

Understanding that technology adoption is a challenge in this domain, we drew on Bezboruah et al. [2014]'s work, which suggests the involvement of staff in identifying existing effective strategies and then leveraging these as models in a design process. Following this procedure, we created several features to address user needs. These included video, image, and audio update interfaces; text and audio chat-style Q&A for family–staff communication; and separate interface views for family members and for staff, each providing detailed care notes. See Figures C.1 and C.2 in the Appendix for examples of artifacts created when designing these interfaces.

5.3.5.2 Iterative Prototyping

To evaluate and iteratively refine our prototypes, we employed the think-aloud method during simulated caregiver scenarios [Ericsson and Simon, 1980]. Initially, we explored a wide range of interfaces through low-fidelity prototypes such as paper sketches, storyboards, and

simple wireframes [Snyder, 2003], sharing these early designs with a subset of caregivers and family members (N=16) for feedback.

Low-Fidelity Prototypes: During scenario-based testing, we presented participants with multiple low-fidelity prototypes and a series of prompts simulating real-life care events that varied in complexity and urgency. Participants were asked to navigate the prototype based on their natural reactions to these prompts, allowing us to observe how their engagement levels fluctuated depending on the scenario. These scenarios were designed to assess how family members would prefer to receive and interact with care-related information under different circumstances. For example, one scenario depicted a routine, non-urgent update about their loved one attending a scheduled group activity, while another described a more critical situation, such as an unexpected change in medication or an incident requiring staff intervention.

High-Fidelity Prototypes: Based on insights from early prototyping, we developed and iteratively improved a single high-fidelity prototype—a mobile-first web application backed by real-time data entry via Firebase. We tested our high-fidelity prototype using the Wizard-of-Oz technique [Dahlbäck et al., 1993], which involved an additional interviewer who actively updated the database in response to user interactions. The web application listened for Firebase updates, creating the impression of a fully dynamic, functional system. To keep our prototyping nimble, we simulated particularly complex UI components as static images that we could update manually and re-upload. Loading bars, which we could quickly trigger on or off, as well as an explanation that this was a new, AI-based system that took time to process feedback, all helped make the lag in updates appear more realistic. As a temporary stand-in for the not-yet-complete backend integration with existing systems, we mocked up a small sample dataset with expected results. Before testing with caregivers at facilities, we tested with care experts from Memcara, organizing feedback as a heuristic evaluation [Nielsen, 1994]. Each walkthrough with caregivers was

conducted by two interviewers who took notes and then led a discussion of the findings in weekly meetings.

Functional Prototype: Having created a high-fidelity specification for FITL that addressed the question of how to design an interactive technology that supports our identified needs and enhances relationship-building to foster a greater sense of engagement for families of PLWD (RQ2), we implemented a fully functional prototype. This involved programming a robust frontend, creating backend integrations with Memcara for passive data collection, and the development of a data processing pipeline.

5.3.6 Delivery (Addressing RQ3)

5.3.6.1 Pilot Testing and Refinements

Prior to the broader deployment of the system, we conducted multiple pilot tests with small groups of staff caregivers and family members of care recipients at two facilities, which had been using the prototype for three weeks by that point. Interview debriefs, surveys, and system usage logs helped us identify bright spots (e.g., families valued the emotionally resonant short videos) as well as friction points (e.g., the AI-generated bullet points needed clearer explanations and citations). We refined the backend data pipelines for video capture and analysis, redacting unrelated or sensitive footage. Through these refinements, we arrived at our final version of FITL.

5.3.6.2 Deployment and Evaluation

Having completed our pilot testing, we deployed FITL at partnering care facilities to evaluate FITL's real-world impacts on family engagement and staff caregiving practices. We were careful to mitigate additional workloads on the staff. Ultimately, we limited our deployment to four of the seven participating facilities due to unanticipated administrative

departures at the excluded sites[2]. We provided short training sessions about using FITL to each site's care staff. Family members received secure login credentials and a video-introduction. Deployment lasted three weeks at each site, enabling us to observe whether families felt more involved in their loved ones' care and whether staff found the system's data-sharing tools efficient or burdensome. We decided on three weeks as an appropriate timeframe because it allowed sufficient time for staff and family members to develop familiarity with the system, engage in multiple interactions, and provide meaningful feedback—while remaining short enough to minimize potential disruption to care routines and staff workload.

To gauge FITL's impact on families' sense of involvement in care as well as other caregivers' practices (RQ3), we used on-site observations to assess how staff integrated the FITL tools into their daily workflow and the effects on care outcomes. My co-authors and I conducted in-depth interviews with participants who could find time outside of work hours (six family members and eight staff caregivers) at the end of the trial to understand shifts in perceptions of communication quality, family involvement, and trust in the caregiving team. Participants who were not willing to participate in interviews cited scheduling challenges as the main reason. Qualitative data collected during the deployment phase (interviews and on-site observations) was thematically analyzed by three coders following the same approach discussed in Section 5.3.4.

5.4 Findings

I begin with insights from the Discovery phase (RQ1), which surfaced a range of interwoven, information-related challenges. I then describe how these findings informed the design and development of the prototype (RQ2), before concluding with results from the evalua-

[2]After consulting with the staff in these cases, we decided to discontinue the study due to concerns that interviews and outside observers would distract the remaining staff at a difficult time.

tion of the deployed system and its impact on family engagement and caregiving practices in real-world care environments (RQ3).

In reporting on themes throughout this chapter, I refer to staff caregivers and family members using "Fx-CGy" and "Fx-FMy", respectively, to identify their facility ID (F) and participant ID prefixed to indicate their status as a caregiver (CG) or family member (FM) – e.g., F1-CG1, F1-CG2, F1-FM1, and so on.

5.4.1 RQ1: What are the experiences of families and professional caregivers of care recipients living with dementia in long-term care facilities, particularly with respect to challenges in communication and family involvement?

From the qualitative investigation into the experiences of families and professional caregivers at long-term facilities, two central themes emerged in the analysis: (1) **Staff-Centered Caregiving Needs**, capturing staff experiences in dementia management[3], and (2) **Family Involvement & Information Needs**, highlighting how families engage and seek transparency in care. Dividing the findings into these role-based themes allowed us to surface the distinct—but often overlapping—needs, practices, and constraints faced by each group. This structure helped clarify how breakdowns experienced by one role often ripple into the other group, and enabled us to examine tensions, expectations, and opportunities for improved coordination. In Section 5.4.1.3, I return to the question of how roles are negotiated and evolve, particularly in high-involvement cases.

[3]Beyond what is noted in Tables 5.1 and 5.2, my co-authors and I initially attempted to further distinguish between general caregiving needs and those specific to dementia, but found that dementia shaped nearly every aspect of care—sometimes in subtle ways, and sometimes in more pronounced ones. In most cases, it compounded practices that were already under strain. While this chapter focuses specifically on dementia care, the approaches proposed here may be applicable more broadly, given that dementia represents a particularly challenging use case. I leave it to future work to further examine this distinction.

5.4.1.1 Theme 1: Staff-Centered Caregiving Needs

As discussed in Section 5.2, dementia care adds significant complexity to already demanding caregiving work. In addition to routine responsibilities like hygiene and feeding, staff must respond to unpredictable behaviors, emotional volatility, and needs that change not just daily, but hour to hour. These interactions often depend on subtle, personal knowledge—what calms one resident might distress another—and mistakes in redirection or communication can have serious consequences. Yet this person-specific knowledge is often missing, fragmented, or inaccessible in existing systems.

Staff operate in environments with high turnover, limited continuity, and tools that rarely support the kind of flexible, context-rich care that dementia demands. To understand where current systems break down and where design can help, I organize the findings into three subthemes: **the challenges staff face** in daily dementia care, **the workarounds and informal practices** they use to fill gaps, and the **design constraints** that shape what kinds of interventions are actually feasible on the ground. These subthemes reflect how staff make care work—not just through protocols, but through improvisation, communication, and shared knowledge that often goes undocumented.

Table 5.1: Theme 1: Staff-Centered Caregiving Needs

Subthemes ***(with associated codes)***	**Description**
Staff Challenges ***(rooted in the demands of dementia-specific symptoms)***	
Symptom management support	Managing challenging dementia behaviors
High workload	Heavy demands limit dementia-specific caregiving
Staff burnout and turnover	Turnover disrupts dementia-specific caregiving
Family-facing communication	Communicating what residents can no longer express
Staff Existing Practices & Limitations ***(illustrations indicate ties to design decisions in Section 5.4.2)***	
Briefings on care essentials	 Notes on managing dementia symptoms
Internal staff conversations	 Conversations about care recipient specific info
Archives of records	 Standardized records to meet legal requirements
Staff Design Constraints ***(broadly applicable to care beyond PLWD)***	
Information continuity	Insights must persist across staffing changes
Minimal caregiving disruption	Solutions cannot increase caregiver workload
Workflow integration	Tools must align with existing caregiving practices

Staff Challenges

Symptom Management Support. Staff repeatedly emphasized that effective dementia care depends on knowing each resident's triggers and calming strategies. As F3-CG1 explained, misunderstanding a resident's needs can escalate into false reports of aggression:

> *"Most people think that, as a caregiver, you want to hug them, but [this resident], she would whack you... You've created an incident where she looks like she's violent because [you] are trying to redirect her in a way that is not appropriate."*

Instead, staff described how a simple memory cue could soothe her, such as referencing "The Golden Girls", which the resident watched with her sisters. However, this kind of detail was rarely available to new or rotating staff. As F2-CG2 noted, the lack of centralized, individualized profiles meant staff had to rely on guesswork:

> *"If I was coming in as a new employee, I wish I could click on a name and I could see: this is their family history... these are the activities they're successful at... these are triggers—that if they're gonna be aggressive, don't do these things."*

High Workload. Caregivers reported being stretched thin, with little time to reflect, observe, or document nuanced changes in behavior. As F4-CG1 noted, tools that demand real-time responses are unlikely to succeed:

> *"Any tool that relied on real-time responses from staff would likely fail unless it fit into moments when staff had bandwidth."*

The pace of work often meant documentation was reduced to checkboxes and surface-level entries: "cleaned," "fed," or "agitated," which contributed little to ongoing care quality.

Staff Burnout and Turnover. Participants linked high turnover to a breakdown in continuity of care. When experienced staff left, their knowledge of residents—what worked, what didn't—was rarely captured or transferred. F3-CG2 stressed that staff burnout limits caregivers' ability to go beyond task completion:

> *"The system should give the caregiver a break... so they can do their own self-care and better focus on what they need to do."*

Family-Facing Communication. Staff viewed family communication as essential but resource-intensive. They want to share meaningful updates—especially moments of care recipient joy and progress—but without overwhelming themselves or families. Caregivers

also noted that without regular, proactive updates, families sometimes doubt that care is happening at all. For example, a family member might call and ask a care recipient what they've done that week, only to hear, inaccurately, that they've been alone or locked up, simply because the care recipient doesn't remember. These moments can create unnecessary distress and conflict, even when care has been consistently provided.

Beyond emotional reassurance, staff saw regular updates as an important way to help families understand how their loved one is changing and what kinds of care strategies are now needed to bring out moments of clarity, calm, or connection. These updates help families build more realistic expectations and feel more connected to the evolving care process.

Staff described a range of family communication needs, which varied in intensity and purpose. At the most basic level, caregivers emphasized the importance of lightweight updates—brief notes, photos, or short videos that offer reassurance and visibility into everyday care without placing additional burden on staff. These updates were especially helpful for sharing moments of connection or progress and for coordinating visits based on care recipient rhythms, such as after music therapy when they were most alert. As F3-CG1 explained, "This would be the most advantageous time for visits."

Beyond periodic updates, caregivers also noted the need for a mechanism to support clarifications or follow-up questions from families. These often arise when a care recipient provides an inaccurate account of their experience—such as saying they've been alone or locked up all week, simply due to memory loss. In these cases, caregivers wanted a way to share more detailed context or documentation without being drawn into ongoing back-and-forth messaging.

Finally, for families who are more actively involved—or who are concerned about consistency in care—caregivers described a need for systems that can offer accessible, on-demand evidence of care. This might include activity logs, behavioral notes, or short annotated media clips that show care patterns over time. However, staff cautioned that systems

should avoid overwhelming families with raw or decontextualized data, which could misrepresent the natural fluctuations of dementia or unintentionally erode trust. Across all tiers, caregivers expressed a need for communication tools that are responsive to family needs but aligned with the realities of their workload and the nature of dementia care.

Staff Existing Practices and Limitations

Briefings on Care Essentials. Caregivers often create informal "cheat sheets" with quick tips on managing resident behaviors. These are typically posted on walls, left in binders, or shared verbally during shift changes. However, these tools are inconsistently maintained and rarely updated, making them unreliable during staff transitions. F4-CG1 described a failed attempt to share knowledge via a physical notebook:

> *"Let's say the daughter writes in the notebook every day... Then one day she forgets. The next day the housekeeper puts it in a drawer. The next thing you know, the notebook's under clothing. The system just fails."*

Reviewing Records. Caregivers sometimes review resident charts, intake forms, or digital notes before a shift—especially when covering a new wing or unfamiliar resident. These records, often embedded in electronic health record systems, are meant to provide a foundational understanding of the resident's history. However, staff repeatedly reported these systems are difficult to navigate, poorly maintained, and rarely helpful for day-to-day decision-making in dementia care.

Many caregivers described situations in which missing or buried information—such as behavioral triggers, preferred redirection strategies, or personal history—led to avoidable distress. F3-CG1 shared an example where a resident with an aversion to touch was approached by a new staff member unaware of her sensitivities, provoking a negative reaction and resulting in a mischaracterization of the resident's behavior. Without timely access to context, she explained, staff risk escalating situations that could have been prevented:

"You've created an incident where she looks like she's violent because [you] are trying to redirect her in a way that is not appropriate."

Even when resident-specific context is captured somewhere in the record, it is often buried within lengthy, compliance-driven documentation that caregivers don't have time to sort through. This mismatch between what's documented and what's actionable can have serious consequences—as seen in incidents where staff unknowingly repeated harmful actions due to missing or inaccessible information.

Beyond its importance in moments of direct care, caregivers also highlighted how personalized context supports broader responsibilities like activity planning and family engagement. F1-CG1, a programs manager, explained how reviewing resident activity logs helped her identify patterns in engagement and tailor programming more effectively:

"This person is always saying 'No'... but we have to look at why they're saying 'No'. They just like to keep to themselves? Or is it because we're not offering them anything they want to be doing?"

She noted that this type of insight is also helpful during family meetings, when staff need to explain what has been tried, what has worked, and what adjustments are being made. Yet this information is typically scattered across paper logs, mental notes, and disconnected digital systems—making it difficult to retrieve when needed most.

Staff Conversations. Caregivers rely heavily on brief verbal exchanges—often in hallways or during shift transitions—to share timely observations and coordinate care. These informal handoffs are a vital part of how teams adapt to changing resident needs in real time. However, caregivers noted that these conversations are often unstructured, easily forgotten, and highly dependent on individual memory and follow-through. Participant F4-CG1 explained how communication is often constrained:

"There is not enough time to have a sit down every day... I could say to the social worker, 'I realized today that Mrs. Jones needs new clothes,' and she

says, 'Okay, I'll call the daughter.' " However, if the social worker forgets, this communication also fails.

While these conversations are essential to keeping care running, staff repeatedly emphasized their limitations. Important details—both short-term updates like a clothing need or changes in mood, and longer-term insights such as a resident's response to certain activities or preferred redirection strategies—can easily slip through the cracks if not captured or reinforced elsewhere. As one caregiver noted, this kind of person-specific knowledge is especially vulnerable to loss during shift changes or when staff rotate across units.

Caregivers expressed frustration at the fragmented nature of current communication practices, with information spread across notebooks, whiteboards, casual conversations, and digital systems that rarely sync. As F1-CG1 put it:

"It's all in different pockets right now, not all on one platform."

Together, these reflections point to the need for tools that preserve the immediacy and informality of verbal exchange while ensuring important information is captured, retrievable, and shared across roles and time.

Staff Design Constraints

Information Continuity. Caregivers emphasized that resident-specific insights—such as triggers, calming techniques, and care preferences—must persist across staffing changes and organizational churn. These insights are often hard-won and critical for avoiding repeated mistakes. F7-CG1 shared a tragic incident in which a care recipient, due to stroke-related language confusion, repeatedly said "yes" when she meant "no." Without access to this crucial detail, a rotating staff member misunderstood her agreement as consent, leading to serious physical harm and legal consequences. Although this information existed in the digital record, it was not surfaced to the caregiver in a timely or accessible way.

Minimal Caregiving Disruption. Any new tool or process must respect the constraints of fast-paced care environments. Caregivers emphasized that systems requiring extra effort, multitasking, or cognitive load are unlikely to be adopted. As F2-CG1 put it, the system should "help us—not slow us down."

Workflow Integration. Designs should extend—not replace—existing informal practices like resident cheat sheets, conversations between caregivers, and deep dives into records. Tools must adapt to the rhythms of care work and slot into existing touchpoints without demanding new ones.

Reflection on Staff's Dementia Caregiving Needs One of the clearest takeaways across the interviews was that care recipient-specific knowledge—such as behavioral triggers, calming strategies, and preferred routines—is essential to high-quality dementia care. Yet caregivers consistently noted that this information is fragile: it often resides in the heads of long-tenured staff, scribbled on informal notes, or buried deep in compliance-oriented records that are hard to access. When staff rotate or leave, that knowledge is frequently lost.

This fragility is not merely inconvenient—it can lead to real harm. F7-CG1 shared a tragic example of a care recipient who, due to progressing dementia symptoms, consistently said "yes" when she meant "no." This detail had been documented, but it was never surfaced in a way that new staff could easily access or understand. As a result, a rotating caregiver misinterpreted the resident's responses, leading to serious injury and legal consequences. Although the system had recorded this insight, it failed at the point of care.

This incident reflects a broader pattern identified in Theme 1: *caregivers do complex, relational, high-stakes work in systems that do not support the continuity this work demands.* Such failures are compounded by high workloads, rapid staff turnover, and fragmented information-sharing practices. As noted in previous sections, information gets

passed in hallway conversations, written on cheat sheets, or tucked away in binders—none of which reliably persist across time or roles.

Designing for **information continuity** requires more than just building digital records. It requires creating mechanisms that surface what matters, when it matters—without burdening staff with extra documentation or workflows. As caregivers made clear throughout the interviews, the goal is not just more information, but more accessible, actionable, and enduring knowledge that supports care over time. Supporting continuity, then, means designing within real-world constraints: tools must fit into tight workflows, preserve the nuance of person-centered care, and work across rotating and unevenly trained teams.

5.4.1.2 Theme 2: Family Involvement & Information Needs

Families remain emotionally and practically involved even after their loved ones enter residential care. This role becomes more complex as dementia progresses and residents lose the ability to self-advocate. Families depend on staff for regular updates but often feel excluded or under-informed about daily experiences.

Table 5.2: Theme 2: Family Involvement & Information Needs

Subthemes *(with associated codes)*	**Description**
Family Challenges *(rooted in the demands of dementia-specific symptoms)*	
Limited visibility into daily care	Communicating what a loved one can no longer express
Eroded trust in care quality	Lacking trust in staff's dementia caregiving
Emotional responsibility without agency	Feeling responsible but unable to influence dementia care
Families' Existing Practices & Limitations *(repeated illustrations from Theme 1 indicate similarity, see Figure 5.2)*	
Informal updates	 Occasional updates on care recipient's status
Conversations with staff	 Follow-up conversations about care updates
Care conferences	 Scheduled meetings for care review and discussion
Family feedback	 Families share stories and care tips
Family Design Constraints *(broadly applicable to care beyond PLWD)*	
Family-staff coordination	Family knowledge integration cannot burden staff
Privacy and compliance clarity	Information sharing must navigate privacy regulations

Family Challenges

Limited Visibility into Daily Care. In dementia care, where residents often cannot reliably self-report, families rely on staff to understand how care is unfolding. Many described frustration with inconsistent updates—typically shared through sporadic emails, calls, or in-person conversations—leaving them without a clear sense of whether care reflected their loved one's needs. As F4-FM6 shared, *"I want to feel confident that my loved one's care reflects her needs and preferences, even when I'm not there."*

Families expressed a desire for structured, tailored communication: regular briefings, the ability to ask clarifying questions, and access to supporting details when needed. These moments helped them stay engaged and aligned with care goals, especially when routines

or behavioral responses changed. At the same time, families recognized that time constraints, documentation challenges, and staffing shortages made consistent communication difficult. Rather than constant contact, they sought systems that supported flexible, layered information sharing without adding burden to caregivers—helping them feel informed, involved, and emotionally reassured.

Eroded Trust in Care Quality. Limited visibility often fed into a broader erosion of trust. In dementia care settings, where families cannot rely on the care recipient's memory or verbal accounts, trust in the care team becomes especially central. Families described a lingering sense of unease about whether care was being delivered consistently and attentively. These concerns weren't always tied to specific events but accumulated over time—through inconsistent staff interactions, lack of follow-through, or mismatches between what families observed and what they were told. As F5-FM2 noted,

> *"The care my mother receives has improved somewhat, but inconsistencies remain. Some staff members consistently provide excellent care, while others seem less engaged, creating significant uncertainty in the quality of care".*

Trust, families emphasized, was not solely about visibility. It was shaped by the tone and responsiveness of communication, a sense that staff understood and respected the care recipient's needs, and prior experiences of either feeling included or left out. In dementia care, these relational aspects are heightened: because care recipients may struggle to self-advocate, families must rely on staff not only for care delivery but also for interpretation and representation of their loved one's needs. Without systems that support consistent, empathetic communication, trust was easily strained.

Emotional Responsibility Without Agency. Many families described the emotional weight of feeling responsible for their loved one's wellbeing—especially in dementia care,

where the care recipient may be unable to self-advocate—while lacking meaningful ways to influence care. As one family member put it,

> *"I still consider myself her primary caregiver... I help manage anything that needs to be managed"* (F4-FM1).

Despite this sense of responsibility, families often felt excluded from decision-making. They described offering input or asking questions, only to see little follow-through and repeated incidents of avoidable caregiving errors. This disconnect left them feeling anxious and powerless, caught between wanting to help and lacking the tools or opportunities to do so.

Existing Practices and Limitations

Informal Dementia Updates. Some families described receiving occasional updates via email or SMS, sometimes accompanied by photos or video clips. These moments were deeply meaningful—such as a short video of a care recipient singing along during a group activity. F5-CG1 shared an example of calling a family member after a previously nonverbal care recipient began singing along during a group activity:

> *"I called the family, and I just said… 'You know that your loved one actually started singing along.' And they were so happy and touched that I reached out to them for something like that."*

However, these kinds of updates were more aspirational than routine. Caregivers explained that recording and sharing media was difficult to scale. Capturing a moment of lucidity in dementia care often requires timing and attentiveness that conflict with the demands of active caregiving. Attempting to record during care may disrupt the moment itself, while recording continuously and reviewing footage to find shareable clips is far too time-intensive. As a result, media updates were generally left to administrative or managerial staff, and occurred infrequently. Caregivers also expressed caution about taking

initiative in sharing updates due to unclear boundaries around privacy and communication protocols. This left families wishing for more regular insight, but unclear on whether or how it might realistically be provided.

Conversations with Staff. Follow-up questions or clarification requests were often addressed verbally during visits or by calling the facility, but these exchanges were unpredictable. Families valued being able to check in and ask specific questions about observed changes or strategies in use, yet noted that the effectiveness of these conversations hinged on who was available and whether the staff member had relevant knowledge to share.

Scheduled Care Conferences. Most families participated in periodic care meetings (often quarterly) where staff shared updates on the care recipient's condition, medications, and long-term care planning. One key strength of these conferences was that staff prepared in advance by reviewing and organizing more detailed records, which meant families could access summaries or explanations that were otherwise difficult to obtain. Families appreciated having this structured opportunity to ask questions and hear directly from multiple care team members. However, because these insights were not available on an ongoing basis, families still felt disconnected between meetings—especially when trying to understand day-to-day care dynamics.

Family Design Constraints

Family–Staff Coordination. Designs should enable families to contribute knowledge and stay informed in ways that are lightweight, asynchronous, and minimally disruptive to staff workflows. Tools that require extensive time or attention from caregivers are unlikely to be used consistently and may add to existing burdens.

Privacy and Compliance Clarity. Communication systems must clearly embed privacy and compliance guidelines, making expectations transparent for both staff and families.

Doing so can reduce staff hesitation around information sharing and help families understand what types of updates are appropriate and available.

5.4.1.3 Integrative Reflections: Entangled Roles and Informational Needs

Dementia care amplifies general caregiving challenges by introducing deep person-dependence, emotional volatility, and informational asymmetry. While staff attempt to manage this through informal workarounds, and families attempt to stay connected via ad hoc inquiries, the absence of structured, persistent information-sharing mechanisms produces gaps.

Both staff and families expressed a desire for shared, contextual knowledge that is easy to access and update. A family member put it plainly: *"I want to feel confident that my loved one's care reflects her needs and preferences, even when I'm not there"* (F4-FM6). To meet that goal, dementia care systems must prioritize continuity, transparency, and communication—without placing additional strain on already overburdened caregivers.

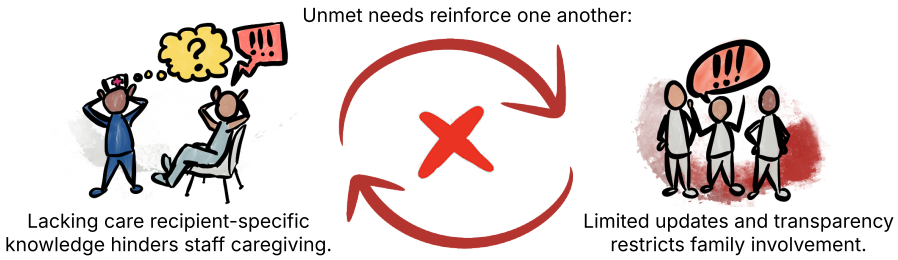


Figure 5.1: Staff and family needs are deeply intertwined—when one side's needs go unmet, it often increases pressure on the other. *All illustrations in this paper are the first author's hand-drawn interpretations of the findings.*

The analysis of staff and family roles highlights a key tension: although their responsibilities and proximity to care differ, their informational needs often converge. This is especially evident when family members become more involved—due to a change in condition, proximity, or advocacy role—and begin seeking not just updates, but contextualized knowledge that helps them interpret what is happening and respond accordingly.

While my co-authors and I began this analysis expecting to delineate staff and family needs as distinct, we found that their roles are more entangled than discrete. Staff may be oriented around documentation, compliance, and clinical decision-making, but they also serve as relational intermediaries for families. Likewise, even families who are not physically present still act as caregivers—albeit asynchronously—by contributing context, asking for clarification, and offering insights into the resident's history and preferences.

This convergence does not mean their needs are the same, but rather that they intersect in ways that call for shared tools, not siloed ones. The challenge is to accommodate different forms of presence, authority, and emotional investment without forcing alignment. Families and staff do not simply need access to the same data—they need tools that reflect the different ways they make sense of and act on that data.

This finding guided the system design toward supporting both differentiation and overlap: tools that provide shared visibility while respecting role-based boundaries, and that accommodate the fluid, often shifting nature of caregiving responsibilities over time.

Zooming out, my co-authors and I also came to understand these needs as part of a broader care ecosystem—one that extends beyond the facility to include hospitals, outside clinicians, and administrative partners. Within this larger system, staff and family needs are not isolated; as shown in Figure 5.1, they are tightly interwoven. Breakdowns in communication or coordination ripple outward. When staff lack insight into a resident's personal history, preferences, or behavior patterns, they may struggle to tailor care in ways that feel meaningful. In turn, families who perceive care as impersonal or inconsistent often escalate concerns, leading to repeated requests or emotional tension—further increasing pressure on already overstretched staff.

This creates a cycle where everyone is trying to do their best, but in the absence of the right information, they end up working at cross-purposes. I see an opportunity to break this cycle through what I call care transparency—the consistent, two-way sharing of contextual and behavioral insights between staff and families. In dementia care, where residents

often cannot advocate for themselves, this shared information becomes the foundation for personalized, responsive treatment.

Through bringing families into an **information-sharing loop** within the broader care ecosystem, staff could access valuable context that supports more personalized care, while families would gain clarity and confidence in how care decisions are made. Designing for this kind of transparency is less about monitoring and more about building shared understanding—helping both sides align around the unique needs of the person at the center.

5.4.2 RQ2: How might we design and build an interactive tool that addresses the needs of family members and caregivers and enhances relationship-building to foster a greater sense of engagement for families of PLWD?

I designed FITL as a single, unified system that supports both caregiving staff and family members of PLWD. While the RQ1 analysis surfaced distinct needs and constraints for each group, I made a deliberate decision to avoid bifurcating the system into separate tools. Fragmentation of this kind is not only technically burdensome but structurally misaligned with the nature of dementia care, where family and staff needs are deeply interdependent and unfold within a shared informational and relational ecology.

Instead, I developed a shared infrastructure that offers differentiated access and interaction modes, while preserving informational continuity and responsiveness across both roles. By designing for overlap rather than separation, I was able to create a system that works within the real constraints of care environments—high workload, legal regulation, interpersonal dynamics—while enabling new modes of coordination that extend beyond what either group can accomplish independently.

The following two tables outline the specific design decisions made in response to the themes that emerged in RQ1. In each case, I sought to emulate existing, situated practices already in use by staff or family members, while addressing their well-documented breakdowns: issues of scalability, information persistence, interpretability, and coordination friction. These design decisions are not speculative proposals, but grounded responses to the informal scaffolding and infrastructural gaps described by the participants. Each decision is shown alongside icons that reference earlier practices documented in the thematic analysis (see Section 5.4.1), emphasizing the continuity between what caregivers already do and what FITL aims to support and scale.

5.4.2.1 Staff-Focused Needs and Design Decisions

Informed by recurring staff challenges around **Symptom management support**, **High workload**, **Staff burnout and turnover**, and **Family-facing communication**, I prioritized continuity, low-burden interaction, and integration with existing workflows. The design intentionally builds on staff routines—like shift notes, informal knowledge handoffs, and basic documentation—while working around the limitations that often render these practices fragile, inaccessible, or unsustainable over time.

Table 5.3: Staff-Facing Design Decisions (Referencing Existing Practices)

Subtheme from RQ1	Design Decision
Staff challenges	
Symptom management support	 Provide structured "Essentials" panels modeled after informal shift notes and hallway exchanges, surfacing person-specific tips often buried in records or lost during turnover.
High workload	 Introduce passive video capture to scale documentation practices without requiring staff to interrupt or alter care routines.
Staff burnout and turnover	 Ensure integration with EHRs to support continuity, drawing on existing archiving practices while reducing friction and duplication.
Family-facing communication	 Create a family-facing Q&A type interface for answering questions and getting feedback, with admin controls for content moderation.
Staff Design Constraints	
Minimal caregiving disruption	 Combine video-based capture with system integration to maintain current task rhythms and avoid parallel documentation burdens.
Workflow integration	 Align the system with existing cognitive artifacts and communication patterns (cheat sheets, informal briefings), allowing it to fold into daily work with minimal adjustment.

These design decisions collectively reflect a strategy of infrastructuring [Islind et al., 2021, Bossen and Markussen, 2010]: rather than introducing new burdens or abstracted

features, I focus on extending and stabilizing the practices staff already rely on. The FITL system seeks to scale knowledge that is otherwise ephemeral and highly individualized—like what soothes a resident or when they are most lucid—into a shared, persistent record that respects caregivers' time and attention.

5.4.2.2 Family-focused Needs and Design Decisions

For families, the key challenges included **Limited Visibility into Daily Care**, **Eroded Trust in Care Quality**, and **Emotional Responsibility Without Agency**. The approach again centered on practices already in use: ad hoc phone calls, conversations with staff, and care conferences. FITL treats these practices as infrastructural elements—carrying both relational and informational weight—but supplements them with mechanisms that offer greater regularity, inspectability, and persistence without increasing the staff workload.

Table 5.4: Family-Facing Design Decisions (Referencing Existing Practices)

Subtheme from RQ1	Design Decision
Family Challenges	
Limited Visibility into Daily Care	  Deliver short, periodic care updates with optional drill-downs, modeled after informal updates and check-ins that families found reassuring but inconsistent.
Eroded Trust in Care Quality	 Provide verifiable notes and annotated video to lend transparency to care practices, referencing the structure of care conferences but accessible asynchronously.
Emotional Responsibility Without Agency	 Enable family-submitted tips and reflections to persist across shifts, formalizing feedback loops that previously disappeared or went unacknowledged.
Family Design Constraints	
Family-staff coordination	 Use a shared interface that recognizes families as care collaborators, coordinating input and visibility through clear role-based permissions.
Privacy and compliance clarity	 Provide transparent, staff-friendly controls over what information is shared and with whom, keeping legal and regulatory concerns manageable.

These design decisions build on existing human infrastructure rather than replacing it. Instead of assuming that family needs can be met through generalized portals or static dashboards, I intentionally designed for layered visibility and emotional resonance, embedding collaboration into the grain of everyday care practices.

Together, Tables 5.3 and 5.4 show how the complex, hands-on realities of caregiving were translated into a system designed to scale without losing sensitivity to those conditions. These design decisions formed the foundation of the system prototype. In response to the fragmented, resource-constrained nature of long-term care settings identified in RQ1, I implemented FITL as a web-based application. This decision was grounded in the need for a platform that could support multi-stakeholder engagement asynchronously, without increasing the complexity or workload for care staff. This approach aligns with prior literature indicating that digital technologies can enhance the efficiency and effectiveness of care coordination by improving information generation, updating, transmission, and exchange along a care recipient's care pathway [Chen et al., 2024]. A web app also allowed me to embed features into an accessible, low-footprint infrastructure—enabling integration with existing documentation practices while offering families continuous visibility without requiring real-time interactions.

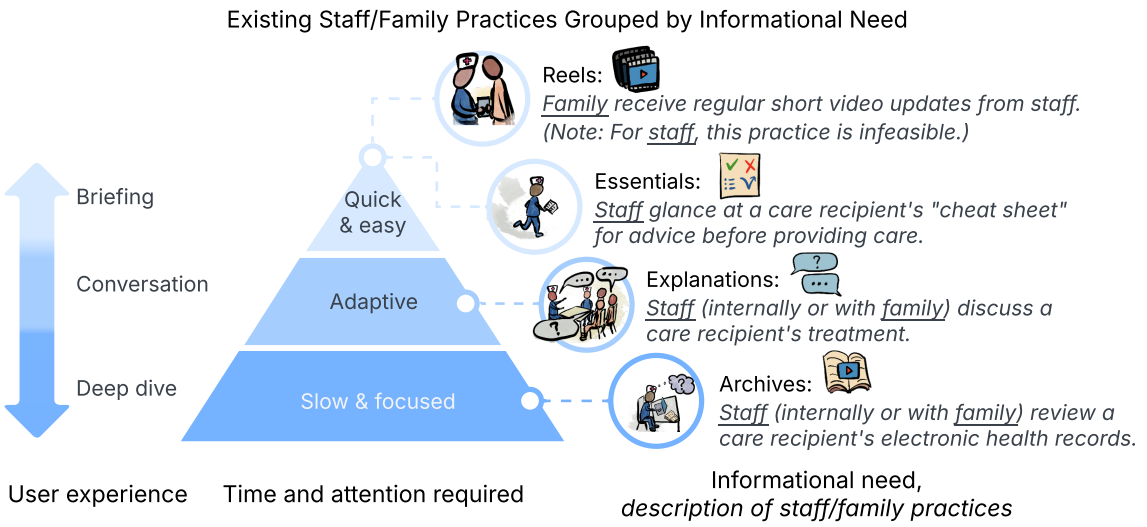


Figure 5.2: Despite distinct roles, family and staff share overlapping information practices. By organizing these practices—including the aspirational model of "Reels"—around the types of informational needs they address, common ground across roles is highlighted—revealing opportunities to design shared tools that support both perspectives while respecting their unique contexts.

The iterative prototyping process focused on designing four linked system views—*Reels*, *Essentials*, *Explanations*, and *Archives*—each corresponding to a common communication practice in RQ1. These views were not designed in isolation but instead reflect an observed spectrum of informational tailoring across staff and family workflows. As shown in Figure 5.2, these views align with varying degrees of time and cognitive effort required from users: Reels and Essentials serve as quick, low-effort entry points for families and caregivers, respectively, while Explanations and Archives support deeper engagement through conversation and review. In the following section, I describe how my co-authors and I developed and refined these components through iterative prototyping and validation with stakeholders—exploring alternative design directions, incorporating feedback across multiple levels of fidelity, and adapting the system to reflect on-the-ground constraints and preferences.

5.4.2.3  Backend Infrastructure: The CareInsights APIs

CareInsights is an interface-agnostic AI service layer with APIs for aggregating, interpreting, and synthesizing heterogeneous care data. CareInsights was intentionally designed as a standalone service with broad reusability in mind. Its development and evaluation are described in detail in Chapter 4; here, the focus is on its role within the FITL system.

CareInsights ingests and organizes data from two primary sources: (1) video recordings of care interactions, and (2) timestamped textual documentation, including shift notes, EHR exports, and informal logs. These inputs are consolidated into a structured, queryable data layer anchored to each care recipient. Video streams are processed using computer vision and speech-to-text pipelines to produce AI-generated annotations approximating caregiver-like assessments of salient moments—such as memory recall, emotional responses, and deviations from baseline behavior. These annotations serve both as inputs for synthesis and as a direct layer of output: underlying assessments (e.g., of memory recall) can be aligned

with transcript timestamps to generate overlays or timelines of interaction patterns, as later illustrated in the *Archives View* (Section C.2.5).

On the retrieval side, CareInsights supports templated, prompt-based queries that return content in formats suited to different stakeholders or interface contexts. A lightweight, LLM-based synthesis layer filters and combines relevant records, producing responses with citations to underlying data entries—such as text excerpts or timestamped video segments. These outputs can be adapted into structured summaries, narrative explanations, or highlight-style compilations.

The backend also accepts user-provided feedback and additions—such as caregiver notes, family observations, or corrections—which are appended to the data layer and dynamically inform future queries. Because the system supports reactive querying, any previously issued or actively "listening" queries are automatically re-evaluated when the underlying data changes. This allows updates—whether prompted by error correction, new observations, or evolving context—to propagate in real time across all relevant views and interfaces using the same data stream.

FITL uses CareInsights across the four primary views, but the CareInsights APIs are interface-agnostic. By cleanly separating UI code from AI/data services, I could prototype diverse frontend representations without modifying core logic. This separation also positions CareInsights as a reusable backend for use cases beyond FITL.

5.4.2.4 Prototyping Process and Insights

Having established the backend infrastructure, my co-authors and I turned to iterative prototyping to explore how CareInsights could support and extend the care practices identified in the initial needfinding. The process included low- and medium-fidelity walkthroughs, scenario-based usability testing, and functional prototypes populated with real-time data from the CareInsights pipeline. These iterations allowed us to refine not only the interface, but also how backend-generated content was interpreted by users in context.

Early prototypes used mocked summaries and static placeholders to explore layout, tone, and content density. As the backend matured, I began connecting real data pipelines to functional UI components. One such feature was the *Essentials* view—a concise, resident-specific summary of behavioral strategies and symptom patterns. When powered by CareInsights, these summaries were well-received by caregivers, who described the content as both accurate and immediately useful, particularly during shift transitions. The summaries reflected patterns of agitation, mood changes, and redirection strategies that closely matched staff observations—suggesting the backend could serve as a reliable cognitive support in practice.

Similarly, family-facing features such as the *Reels* view—a sequence of curated video clips accompanied by generated captions—elicited strong emotional responses. Family members found these compilations both reassuring and informative, often perceiving them as more credible and expressive than written updates. These findings affirmed that the backend could generate content resonating with diverse user needs, from emotional grounding to clinical handoff support.

However, as prototypes became increasingly realistic, new concerns emerged—particularly regarding caregiver comfort with family-facing transparency. Although many staff members initially supported the idea of sharing curated clips with families, several expressed hesitation once the stakes of real deployment became clearer. Legal liability, interpretive ambiguity, and perceived loss of control all became salient, especially when AI-selected video segments were involved. These concerns, absent during early ideation and low-fidelity feedback, underscored the gap between hypothetical and operational acceptance.

To address these concerns, I introduced a series of design controls: media content defaulted to "private," administrative roles were given fine-grained permission settings, and a geofencing option was added to restrict access to certain content unless the viewer was

physically on-site. These controls, configurable by administrator-level staff, were developed in close partnership with site administrators.

These adjustments reflect not only a necessary response to staff concerns around legal risk and content control, but also the value of sustained engagement beyond early design phases. Without long-term input, I might have relied solely on initial enthusiasm—overlooking emergent concerns that surfaced only as the system approached real-world deployment, and risking the creation of a tool staff would ultimately reject.

Another insight shaping the final system design concerned system visibility during routine care. Caregivers valued the **Essentials** summaries in theory but often forgot to consult them during high-stress moments when symptom management was most critical. To address this, I deployed QR-coded signs in resident rooms that, when scanned, opened a privacy-respecting Essentials view tied to the resident's ID. This light-touch intervention created a physical anchor for system use, integrating the tool into existing visual rhythms and care routines.

Table 5.5: Final Pre-Deployment Design Adjustments from Late-Stage Prototyping

Code ***(from RQ1)***	**Updated Insight**	**Final Design Decision**
Family-facing communication	Legal concerns began to outweigh earlier enthusiasm when caregivers realized families would view selected clips.	 Added geofencing and default-private modes for staff-facing clips; gave admins granular control over content visibility.
Workflow integration	Caregivers often forgot to use the system during high-stress moments when it would be most valuable.	 Deployed QR-coded signs in resident rooms to trigger context-specific *Essentials* view; provided quick access without login.

These adjustments reflect the importance of long-term iteration and deep engagement in care settings—where systems must not only function technically but also align with shifting relational and organizational dynamics. In the following subsections, I describe the resulting views of the deployed system and the care practices it supported.

5.4.2.5 Overview of FITL System Views

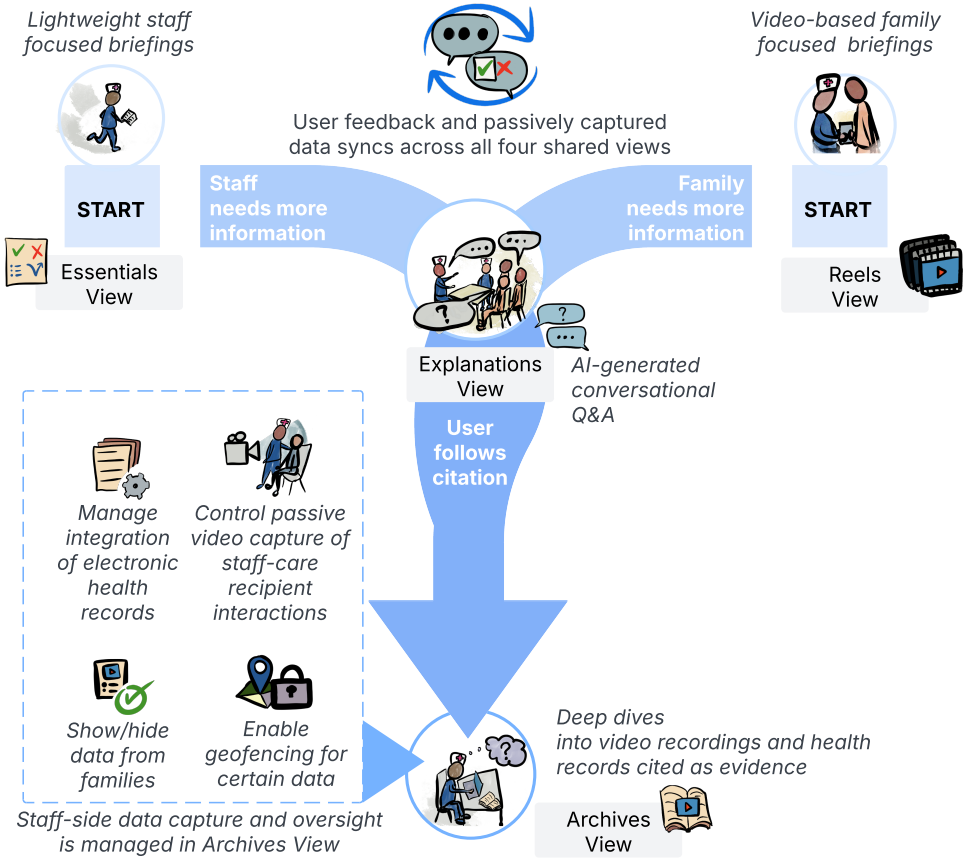


Figure 5.3: Typical user journeys across FITL's four views—Essentials, Reels, Explanations, and Archives. The flow between views reflects how users navigate escalating informational needs in response to emerging care situations, such as changes in a care recipient's dementia-related behaviors. Feedback can be contributed at any stage and is continuously integrated into the shared data layer, enabling all views to dynamically reflect updated understanding, corrections, and newly observed details.

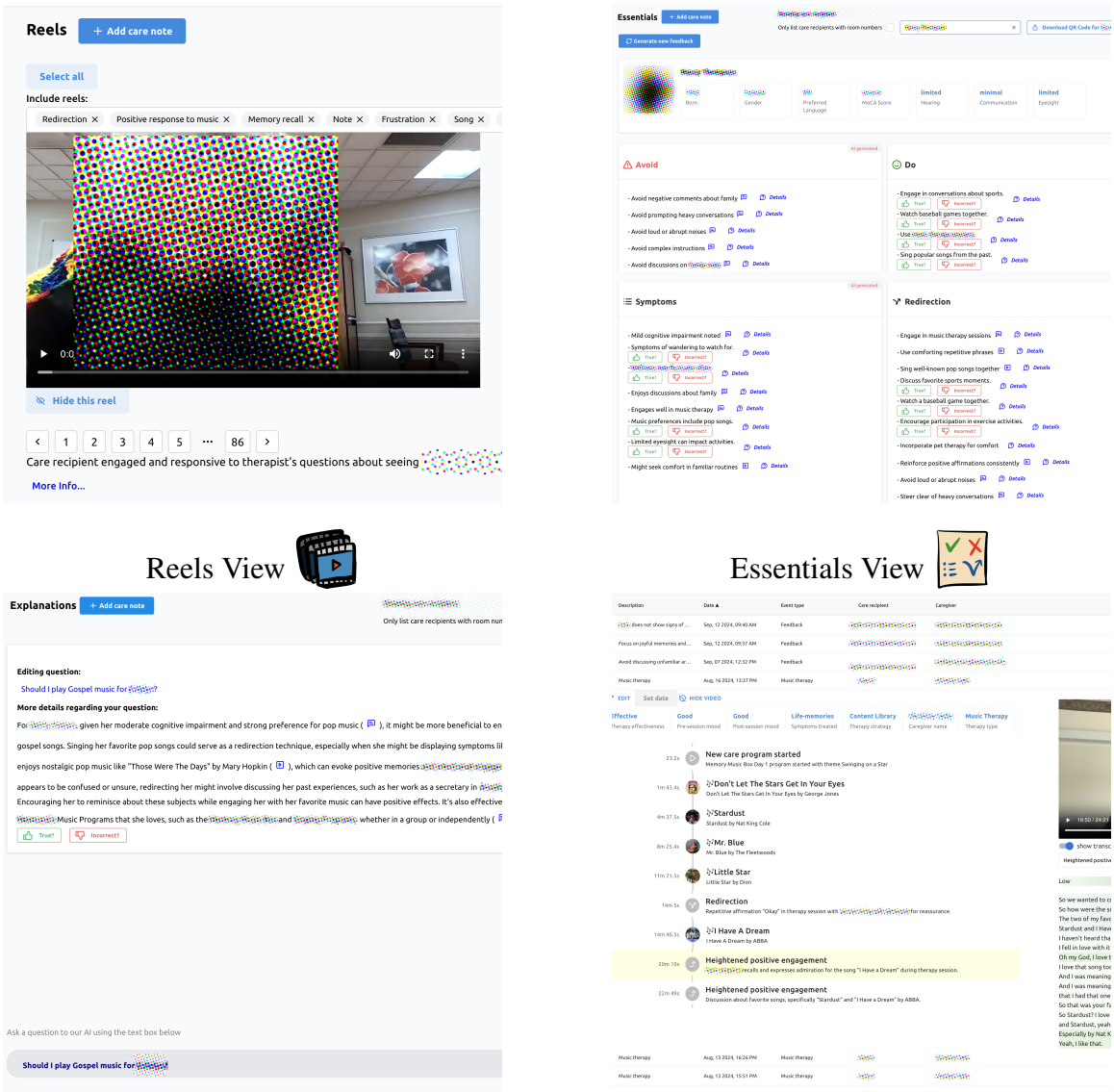


Figure 5.4: Excerpts from screenshots of FITL Views, see Appendix C.2 for a detailed overview.

FITL supports family-staff collaboration through four interconnected views—**Reels**, **Essentials**, **Explanations**, and **Archives**—each designed around existing successful practices that were observed.

Family members commonly began their interaction with FITL in Reels View, where they could see regular updates (on a single care recipient, i.e., their loved one) in the form

of emotionally salient clips that offered a window into recent care. Staff, by contrast, often started with Essentials View to review tailored tips (for any of the participating care recipients) for managing dementia symptoms and supporting resident engagement. These interactions could escalate to Explanations View, which supported clarification and follow-up questions, and further to Archives View, where users could examine the records cited as evidence by the system.

Data flows into FITL through passive video capture, active user feedback, and automated integration with existing EHRs (see Section 5.4.2.3). This source material is used to dynamically update the AI-generated content across Essentials, Explanations, and Reels in real time when needed (e.g., in response to user feedback or a new video recording). The following sections provide a brief overview of each view; **for more detail, see Appendix C.2**.

- **Reels View** offers emotionally resonant video updates, primarily for families, capturing moments from structured therapy sessions with minimal caregiver effort. These clips are algorithmically selected and summarized to foreground narrative or emotional meaning. They help families stay connected and offer a low-friction way to observe care that would otherwise remain invisible.

- **Essentials View** distills personalized care strategies into a concise set of bullet points, modeled after the "cheat sheets" staff sometimes informally maintain. It provides caregivers and families a shared reference for key do's and don'ts, person-specific symptoms, and personalized redirection strategies—each grounded in system-synthesized observations that presented as citations.

- **Explanations View** emulates conversational Q&A with a knowledgeable staff member, offering narrative interpretations in response to open-ended user questions. These were typically questions about care moments—e.g., behavior changes, effective interventions—, trends, or care recipients' memories. Like Essentials

View, Explanations includes citations that a user can follow to Archives View, for instance, to index directly into the exact moment in a longer video recording where the referenced event occurred. Users can provide feedback in the form of open-ended "Care Notes" or quick, "thumbs-up/thumbs-down" interactions that, like all feedback in FITL, cascade throughout the system.

- **Archives View** serves as the full-fidelity interface, enabling users to verify source records, navigate transcripts or higher-level timelines of staff-care recipient interactions, and edit the underlying content. Archives View anchors FITL's interpretive logic and ensures transparency in how summaries, recommendations, and care notes are generated.

Together, these views tailor underlying records to the informational needs of different users.

5.4.3 RQ3: How does the tool impact families' sense of involvement in care as well as other caregivers' practices?

Having discussed user needs and the design rationale behind FITL's features, this section turns to its impact during deployment. It reports how the tool shaped family involvement and staff practices at four long-term care facilities over a three-week period, and also draws on sites that piloted but did not complete deployment. These findings are closely tied to those from RQ1, which documented persistent challenges in dementia care, including high staff workload, limited visibility into daily care for families, fragmented information-sharing practices, and fragile continuity of resident-specific knowledge.

5.4.3.1 Addressing Staff Needs

Integration With Existing Workflows and Perceived Feasibility Staff emphasized that FITL's integration into daily routines—particularly through its use of existing data

sources—was central to its usefulness. Several noted that the tool was available without requiring them to change how they worked. As described in the previous section, staff could engage with the tool as little or as much as they wanted, but appreciated having access to information when needed. F7-CG1 explained that FITL helped him:

> "remember it and pass it on instantly while doing other things… thus saving the time of having a 15-minute conversation about each person."

Others noted that it was especially helpful during handoffs or when interacting with unfamiliar residents, providing quick context that otherwise would have required time-consuming explanations.

The system's unobtrusive design and flexibility made it feel feasible for everyday use, especially in resource-constrained settings. At the same time, however, these same qualities—its accessibility, persistence, and integration into existing workflows—also raised concerns. As described in Section 5.4.3.3, caregivers and administrators expressed unease about privacy, documentation visibility, and institutional risk. In this way, the aspects of FITL that made it practical and sustainable were also linked to ongoing tensions around data sharing and governance.

Supporting Dementia Caregiving Beyond feasibility, caregivers described FITL as a meaningful aid in the complex work of dementia caregiving. The system helped them avoid disruptions in care that stemmed from missing context. For example, F1-CG11 recounted a time when a resident became aggressive because staff unknowingly raised a sensitive topic, while F1-CG12 described a resident becoming distressed when someone else uncovered his food. In her view, FITL,

> "helps prevent some of those accidental slip-ups."

These examples suggest that caregivers saw the tool not as an added burden but as a means to reduce friction and improve care continuity.

By surfacing resident-specific context and preserving institutional memory across shifts and staff turnover, FITL supported the relational, individualized care that dementia requires. Its ability to quietly fill gaps in context—without demanding new workflows—was repeatedly cited as critical to supporting both residents' well-being and staff effectiveness.

5.4.3.2 Addressing Family Needs

Streamlining Communication Through observations and interviews, my co-authors and I found that both family members and care staff felt that FITL enabled better communication through automation features, the platform's data aggregation and real-time capabilities, and search features. Users particularly appreciated how information was available at different levels of granularity, with deep links in generated summaries allowing them to drill down (switching between views) for more detail on demand. One caregiver, for instance, found the bullet points sufficient most of the time but valued having the underlying detail available when it came to redirection. **Family members expressed that this granular view allowed them to witness the nuanced care their loved one received, fostering a greater emotional connection and trust in the caregiving team.** Users also appreciated where FITL flags AI-generated content with a tag (see "AI generated" in Figures C.4 and C.5), distinguishing it from manually entered data by caregivers or family members. One caregiver noted that the system could go even further with this feature,

> "I think it'd be cool to say like, yeah, well, like 7 people found this to be true."

Moments That Matter: Capturing Lucidity and Connection Through Reels During deployment, Reels View offered families a way to engage with aspects of the resident experience that are typically obscured by the nature of dementia. One illustrative case involved F1-CR9, a resident with advanced cognitive impairment. In most recordings, she appeared catatonic or distressed, and care staff had been hesitant to share such footage

with family members, aware that these images could be upsetting. Family members, in turn, had expressed reluctance to view recordings that might reinforce the emotional toll of witnessing a loved one's decline.

However, one clip automatically flagged by the system captured a rare moment of clarity. Toward the end of a therapy session, F1-CR9 turned to the therapist and said,

> "This [therapy] gives me a reason to live."

The statement, brief but lucid, was highly unexpected given her usual presentation. For both the therapist and her family, this reel carried immense emotional weight: it affirmed that the resident could still engage meaningfully and that the therapeutic activities were, at least in this moment, deeply felt.

This reel became especially significant when F1-CR9 passed away shortly after deployment. It remained one of the few captured moments that showed her expressing appreciation or emotion verbally, and it was shared with her family as a form of closure. Staff reflected that this type of insight would have been unlikely to surface without automated detection. Much of the footage collected during the session consisted of silence, minimal movement, or signs of distress—none of which families were inclined to view. In contrast, the reel extracted by FITL allowed families to witness a final moment of lucidity and connection. In this way, the system facilitated an experience that was both emotionally resonant and contextually appropriate, reinforcing the value of AI-augmented visibility into care recipients' inner lives.

In cases like this, surfacing a moment of deep personal meaning often also helped to strengthen families' trust in both the care team and the therapeutic process—cultivating a sense of **solidarity rather than scrutiny**, and affirming that meaningful care was taking place even when it was not immediately visible.

Supporting Clear and Actionable Recommendations for Personalized Care Users found utility in FITL's focus on residents' triggers and preferences, with staff comment-

ing that the available information allowed them to tailor their approach and avoid potential conflicts. For instance, F3-CG5 emphasized the importance of knowing residents' preferences, especially regarding music and conversation topics, stating that for some people it is otherwise difficult to know which memories can be discussed. FITL helped her:

> "avoid controversial subjects that trigger past [negative] experiences."

These adjustments in approach, tailored to individual preferences, were described as fostering a deeper empathy and connection between caregivers and care recipients. This aligns with RQ1 findings showing that emotional investment from families often drives a desire for more actionable and personalized information. A common theme was caregivers reporting that the system offered guidance on interacting with residents—what to avoid and how to redirect effectively. For instance, one caregiver noted:

> "I'm going to look at it in a glance, in the glance, I'll read the avoid section and then I'll read the redirectable section."

Similarly, F3-CG7 praised the efficiency of Essentials View:

> "It comes in handy for people that don't want to sit there and go through everything."

Finally, users appreciated how insights in FITL combined personal narratives with caregiving strategies from nursing literature. F3-CG9 remarked that FITL helped them,

> "get a better snapshot of the patient, not just the medical side,"

reinforcing holistic involvement.

FITL as a Learning Tool Many family members emphasized FITL's value as an educational tool—for themselves and for staff. For example, F5-FM1 noted that misunderstandings stemming from staff inexperience, particularly with comorbid conditions like dementia and aphasia, could undermine care. FITL helped bridge that gap. Family members

described watching recorded sessions in Reels and Archives Views as opportunities to learn what strategies worked well. They appreciated how clips could be filtered and annotated with system-generated cues about resident response, such as memory recall or emotional expression. F4-FM1 explained that these features helped her understand her mother's progression despite not being physically present. **These tools enabled emotional connection and situated learning by observing effective caregiving moments.**

Integration with Existing Systems Throughout the study, users expected the platform to provide useful, integrated information immediately. FITL's authentication and use of Memcara's infrastructure supported this expectation. F1-FM1 appreciated that FITL consolidated care insights that would otherwise remain informal or undocumented:

> "We trade information all the time... but it's kind of an assumed thing... it's not written down anywhere."

By Improving Continuity, FITL Strengthened Staff-family Coordination. Late-stage prototyping revealed that synchronizing event logs from Memcara's interface with video allowed for precise understanding of what stimuli triggered observed responses—surfacing deeper context without increasing staff documentation workload. F4-FM1 proposed additional integrations like a split-screen view to watch stimulus and response together, likening it to "reaction videos" on YouTube. **This reflects a strong desire from family members to feel more immersed in their loved ones' daily care routines, even from afar.**

5.4.3.3 Ethical, Legal, and Privacy Concerns

Privacy Protocols Hindered System Deployment While FITL was well received, privacy concerns—particularly around video sharing—emerged as a barrier. Family members expressed frustration when limited access prevented them from connecting with their loved one's care. **Though reassured that data was protected, family members often found**

the administrative process for account access confusing or unresponsive. The analysis revealed that family involvement was often curtailed not by disinterest but by institutional hurdles—highlighting the tension between privacy safeguards and inclusive care.

Access to Sensitive Health Records As FITL introduced new forms of visibility into care delivery—through video, structured summaries, and contextual records—participants surfaced a range of concerns that reflected not outright resistance, but engagement with the tradeoffs involved. While staff generally appreciated that the tool aligned with their workflows and could be used flexibly, they also pointed to challenges around visibility, including concerns about overexposure, documentation boundaries, and the emotional weight of reviewing care data.

Concerns also surfaced around access to trauma-related records or footage with other residents in view, an issue that was identified but had not been adequately accounted for in the design. **Staff requested more control over what family members could access to avoid unintentional disclosure.** These features attempted to strike a balance between transparency and professional discretion. Yet even as FITL made care more visible, staff sometimes felt emotionally taxed by the idea that every interaction could become reviewable.

Family members, likewise, expressed ambivalence. They valued the opportunity to remain informed, but were often unsure about how much was being shared, who controlled access, and whether what they saw would feel supportive or distressing. Across both groups, concerns around privacy, emotional complexity, and system access highlighted the limitations of existing policies and infrastructure. Ultimately, these responses suggest that efforts to increase transparency in dementia care must be carefully balanced with emotional, relational, and institutional realities, and that even well-aligned tools like FITL can surface new tensions when deployed in complex care settings.

Legal Risks in Recording Some users worried about documentation being misused, but most accepted that video documentation, with appropriate guardrails, could enhance care. **Family members saw these recordings as a valuable source of connection.** Cases were also observed where staff were initially supportive of sharing clips but grew hesitant as deployment neared—underscoring the importance of long-term engagement to surface operational, not just hypothetical, concerns.

Families' Complex Relationship Between Privacy Concerns and Benefits Family members overwhelmingly valued visibility, even when conflicted about data sharing. **Their concerns were not about surveillance but about maintaining their loved one's dignity and control.** For instance, F4-FM6 described the emotional difficulty of "watching from a distance" and wanting to help. As identified in RQ1, these tensions highlight that involvement is not only a matter of access but of identity and advocacy—family members often saw themselves as co-caregivers, despite their distance.

Consent Before Data Collection Finally, users stressed the importance of meaningful, not just legalistic, consent. **Family members wanted clearer ways to engage in decisions about data use and sharing.** F1-FM9 noted:

> "I think it's important that we have a say in what information is shared and with whom."

Staff also expressed that clearer consent processes could help them manage expectations and reduce legal risk, especially as AI-generated records became more prominent in care conversations.

5.4.3.4 Takeaways

The deployment of FITL across four long-term care facilities revealed both the value and complexity of integrating AI-augmented tools into dementia caregiving. FITL improved

families' sense of connection and trust by surfacing emotionally resonant moments and enabling greater visibility into everyday care—often revealing aspects of resident experience that would otherwise remain obscured.

These positive experiences occurred in a care context where new technologies are often rejected if they add even minimal friction. That FITL was not only tolerated but actively used speaks to the urgency and depth of the underlying needs it addressed. Caregivers, who typically operate under intense workload pressures and with significant legal constraints around documentation, chose to adopt the system and allowed data recording that would normally raise red flags. This engagement alone is a meaningful outcome: in environments where emotional and cognitive overload are common, any tool that survives routine use signals feasibility, trust, and unmet need.

Participants voiced important concerns around privacy, emotional labor, and system boundaries—but these concerns were raised not to reject the system, but to make it work. Despite reservations, caregivers and families alike expressed a strong desire to continue using FITL. This willingness to persist reflected a shared recognition that FITL provided a form of visibility and continuity not otherwise available through existing workflows.

Together, these findings point to the need for privacy and access models that reflect the lived constraints of caregiving environments. Participants were not opposed to visibility, but emphasized that it must be implemented with attention to clinical roles, family expectations, and the emotional weight of encountering care data. Designing for privacy in this context involves more than technical safeguards—it requires supporting the relational and operational infrastructures that make data meaningful and manageable.

In the following discussion, I consider how these findings extend current understandings of family-centered care and CSCW system design, highlighting how technologies like FITL can reinforce person-centered practice while navigating the emotional, institutional, and relational dynamics of long-term care.

5.5 Discussion

This chapter introduces **Family In The Loop**, a digital platform that augments dementia caregivers' ability to provide person-centered, family-framed care by supporting knowledge synthesis and communication between staff caregivers internally and with family members. In this approach, I leverage prior literature on family involvement and the discovery and definition process described in Section 5.3 to deeply understand the complex needs of different users. A central insight from RQ1 was that families were expected to trust care processes they could not see, influence, or fully understand—despite possessing critical knowledge that could enhance resident wellbeing. As staff faced systemic constraints, family insights were often lost in the churn of shift work, inconsistent documentation, and reactive caregiving. These breakdowns disrupted the continuity of care and undermined the mutual trust necessary for person-centered practice. FITL was designed to respond to this challenge by making everyday care more visible. A continuous stream of captured therapy videos, along with integrated EHR data, served as passive inputs that allowed the system to surface meaningful moments and improve its insights over time—even without direct user input. This sustained visibility encouraged families and staff to return, reflect, and learn not only from what the system highlighted, but from one another, creating a feedback loop of shared understanding and asynchronous support.

The deployment findings (RQ3) show that FITL was adopted because it aligned with existing workflows and addressed a gap that staff and families both saw as urgent. Staff used the system despite concerns about increased visibility, recognizing that it helped them deliver care more effectively. Families felt more connected and reassured when they had access to curated updates, rather than having to seek out or interpret fragmented information on their own.

5.5.1 Envisioning a Data-driven, Person-centered Care Ecosystem

Stepping back, I position this work as a step toward a future in which data-driven, person-centered dementia care is routine. In that future, AI-enabled tools and adaptive platforms synthesize longitudinal, multimodal data to reduce routine burdens and surface targeted, context-aware care suggestions. Achieving this requires moving from today's data silos and fragmented practices to collaborative, transparent family–staff interactions supported by adaptive tooling. As mediators, digital agents can streamline coordination, open dependable communication channels, and strengthen trust, helping families stay engaged and informed. Integrated in this way, AI can support caregivers in acting on a rich, shared understanding of each care recipient's needs, history, and evolving care plans. The result is not only greater efficiency and improved outcomes for PLWD, but also healthier family–staff relationships and a more sustainable, rewarding caregiving experience.

While FITL improved visibility into the daily experiences of individual residents, the deployment also revealed its limits. The system helped staff and families stay aligned, surfacing meaningful insights in the flow of care—but person-centered care does not begin or end with documentation. Caregivers were often eager to provide more individualized support, yet they worked within systems that prioritized efficiency and risk avoidance. Overmedication remained common, and time for tailored engagement was limited. Even the most motivated caregiver could only do so much. At a certain point, person-centered care becomes a matter of resource allocation: deciding which residents receive which forms of support, and when. These are not just personal or clinical choices—they are shaped by the structural design of long-term care.

Administrators and program coordinators, who manage staffing, programming, and regulatory compliance, were not the intended users of FITL. And while the system offered rich insights into individual needs, it lacked a mechanism for translating those insights into action at the institutional level. Supporting person-centered care at scale requires more than visibility. It demands tools that can synthesize patterns across residents and help decision-

makers weigh those needs against legal, financial, and operational constraints. FITL was not designed for that role—but its limitations help clarify what kinds of systems are needed next.

To meet the kinds of institutional needs that tools like FITL cannot address, I outline a conceptual architecture for a **data-driven, person-centered care ecosystem**. This ecosystem moves beyond supporting individual care decisions to helping facilities plan, allocate resources, and advocate for care practices at scale. Rather than relying on a single system, it brings together multiple layers of technology and insight, each serving a different role in the care environment:

- **The Observational Layer**, responsible for continuous data capture. This includes ambient sensors, mobile tools, wearables, and smart glasses that passively record care interactions, resident behaviors, and environmental context. Such passive capture raises privacy, legal, and ethical considerations [Mujirishvili et al., 2023].

- **The Interpretive Layer**, which transforms raw data into actionable insight. AI agents in this layer, like FITL, detect patterns, surface relevant updates, and tailor information to stakeholders in real time.

- **The Strategic Layer**, where institution-wide trends are aggregated and analyzed. Meta-level systems here support planning, resource allocation, and policy decisions by making the needs and preferences of residents visible at scale.

This layered architecture offers one way of approaching the challenge of scaling person-centered care beyond individual interactions. Rather than expecting a single system to serve all users, it distributes the work of interpretation and coordination across levels, mirroring how decisions are already negotiated in long-term care environments. Observations captured at the point of care—through video, sensors, or routine documentation—do not speak for themselves. To become useful, they must first be processed and framed in ways that are accessible and relevant to the people who rely on them. Families, for instance, are

not in a position to interpret raw data or navigate unstructured streams of updates. They require information that has already been filtered, contextualized, and made emotionally meaningful. For caregivers too, who are pressed for time, the value of observation lies not in its volume, but in its relevance. Without this initial translation, even rich data can go unused. As was heard in the deployment, caregivers often do not have time to watch a full one-minute video, let alone sift through hours of recorded content or interpret sequences of raw metrics. Without tools to extract and present meaningful moments, this information can be more overwhelming than helpful.

The interpretive layer is designed to support this kind of ongoing sensemaking. It enables asynchronous, low-friction ways of surfacing and sharing care-relevant information, helping stakeholders learn from one another even when they are not directly in conversation. For families, this might mean receiving short, emotionally resonant updates that connect them to their loved one's daily life. For caregivers, it may involve viewing brief contextual notes or summaries that help avoid unnecessary disruptions or repeat mistakes. These interactions—though lightweight—begin to accumulate meaning. Over time, they form a kind of organizational memory that makes patterns and concerns more legible across shifts, roles, and time.

Building on this foundation, the strategic layer supports a broader form of learning. Rather than responding to individual needs or events, this layer looks for recurring signals across residents and across time. Its role is to connect the lived experiences surfaced by the interpretive layer with the planning and decision-making practices of institutional leaders. This includes decisions about staffing models, programming priorities, and long-term investments in care practices. Critically, it is not about removing human judgment but about shaping decision contexts that are more informed by what happens on the ground. In doing so, the strategic layer can help facilities better align their operations with the values of person-centered care—without requiring every individual caregiver or family member to advocate alone.

In proposing this architecture, I draw from principles in multi-agent systems, where higher-order agents coordinate decisions by integrating input from task-specific subagents [Leonard et al., 2024]. While this layered approach is well established in computational domains, its application in dementia care remains limited. Coordination in these settings must account for not only clinical complexity, but also emotional, relational, and infrastructural dynamics that are difficult to model. Recent work has explored the role of AI in diagnostic tasks for dementia [Andargoli et al., 2024], but less attention has been given to how such tools might support broader coordination across roles, needs, and time. By outlining this conceptual architecture, I aim to sketch a direction for future systems that move from supporting individual users toward enabling collective action in person-centered care.

5.5.2 Future Work

Realizing this vision of layered coordination will require sustained inquiry and innovation across multiple fronts. Several stakeholder roles—such as program coordinators and case managers—were not directly supported in the initial deployment, but play critical roles in shaping care continuity and team communication. A future expansion of FITL could better integrate these roles, enabling cross-role coordination and decision support at broader organizational levels.

Long-term care systems will benefit from features that track resident health not only in terms of vital metrics, but also narrative-based trends and emotional well-being. Future systems might support the capture and interpretation of longitudinal data about resident responsiveness, sleep patterns, mood fluctuations, or mobility shifts—helping to identify small changes that matter over time.

I see an opportunity to rethink what constitutes meaningful documentation in dementia care. In the field study, my co-authors and I found that conventional formats (e.g., structured EHR fields, cognitive assessments like the MoCA, and checklists of Activities of

Daily Living (ADLs)) fall short in capturing the nuanced and often nonlinear experiences of people living with dementia. These measures are ill-equipped to reflect the emotional depth, narrative texture, or momentary lucidity that can be profoundly meaningful to care recipients (see Section 5.4.3.2), families, and care staff alike.

Without structured ways to account for these human-centered phenomena, the value of person-centered practices risks being discounted in institutional settings. One facility administrator who supported the study, for example, remarked that it was "not worth it" to provide music-based therapy to a particular non-verbal resident. Yet through FITL, subtle but consistent indicators of increased lucidity over repeated sessions were observed—insights that might otherwise have gone unnoticed[4]. While assessing the clinical efficacy of specific therapeutic practices (i.e., those provided by Memcara) was beyond the scope of FITL, the ability to surface and contextualize such moments points to a broader design opportunity. These glimpses of potential therapeutic value show the importance of designing tools that can capture and elevate moments that matter, even when they fall outside conventional documentation or evaluation frameworks.

Future systems might foreground narrative and relational context—what a moment meant, how it was experienced, and how it might guide care going forward. AI can support this shift by detecting and connecting meaningful micro-interactions into richer, evolving storylines that better reflect the lived complexity of dementia.

The ecosystem must also support not only those directly involved in care, but those on the periphery. FITL's utility could be expanded for family members who are unable to visit regularly, through asynchronous communication, voice-based updates, or adaptive storytelling mechanisms that convey key moments without overwhelming detail. This same infrastructure could support augmented reality or wearable technologies to deliver just-in-time information to caregivers on the floor. These directions suggest focusing less on data volume and more on structuring data to support understanding, continuity, and dignity. The

[4]Unfortunately, staff caregivers told us, this story represents a common occurrence.

path forward will require tools that operate across tiers of care: from moment-to-moment support to long-term system transformation.

5.5.3 Limitations

Several limits bound what this study can claim. The deployment ran for three weeks at four of the seven partner facilities, which is long enough to observe uptake and early practice change but not long enough to speak to sustained engagement or to whether families and staff would keep using FITL once the research team withdrew. The evidence for RQ3 comes from exit interviews with six family members and eight staff caregivers together with on-site observation, so the findings are participants' accounts of their own experience rather than measured change; unlike the CareInsights study, no workload or engagement instrument was administered, and the reports that FITL added no burden are perceptions rather than measurements. Those interviews were also limited to participants who could find time outside of work hours, which selects for the people least squeezed by the conditions the system was meant to relieve. Participation was voluntary at facilities that had already agreed to a two-year collaboration, which further selects for staff and families disposed to engage with a new tool. I designed, built, and deployed the system and also collected and analyzed much of the data, so the same person shaped both what the tool did and what counted as evidence that it worked. Finally, the Care Notes that families could send back to staff remain a designed capability rather than a demonstrated influence on practice, since these findings record no instance of one being submitted and acted upon.

5.5.4 Navigating Tradeoffs in Scaling Care Systems

While the deployment surfaced a compelling vision, it also exposed a set of persistent tensions that must be more fully reckoned with as systems like FITL scale. Even in a limited deployment, tradeoffs between automation and manual oversight, and between transparency and privacy, were encountered. Staff appreciated the reduced burden of documen-

tation but expressed concern about how their work might be interpreted. Families valued increased visibility into care, but at times felt overwhelmed or unsure how to engage with what they saw. These frictions reveal deeper questions of trust, discretion, and control—questions that become only more complex as care systems grow more data-driven.

These are not problems with definitive solutions, but ongoing negotiations. Designing for scale means designing for stewardship—supporting continuous calibration between institutional needs and the lived realities of caregiving. In this light, the tensions observed are not side effects of innovation but core design considerations for any future ecosystem. FITL demonstrates that AI can be meaningfully embedded in the everyday practices and relationships that shape care. However, realizing its potential at scale will require more than technical refinement. It calls for sustained attention to the human, organizational, and ethical structures in which these tools operate. The opportunity ahead is not just to expand what care systems can observe, but to strengthen what they are able to understand and sustain.

5.6 Conclusion

Through an iterative design and development process, I created and evaluated Family In The Loop, an AI-driven web application to support person-centered dementia caregiving. Building on prior studies centered on caregiving staff, this work explores a new, often-overlooked direction: empowering care-recipients' families through a dedicated, family-focused portal that is deeply integrated into existing caregiver workflows and existing EHR systems. My work seeks to improve the understanding of how technology can support information sharing and synthesis between families and caregivers. I follow a human-centered, collaborative design approach that is focused on deeply understanding the needs of facility staff, PLWD, and family members. I evaluate my final prototype through a deployment study and present rich qualitative findings from in-person observations and inter-

views. My findings suggest promising improvements in family involvement and trust, and in cross-role communication, which in turn seem to bolster the joint capacity of families and staff in long-term care—as a complex adaptive system—to provide more person-centered care. Finally, I reflect on the limitations of current systems and propose a layered view of data-driven, person-centered care as a path forward. To foster care ecosystems that are both data-rich and deeply human-centered, I offer this framework as a roadmap for AI-enabled systems that connect frontline caregiving with institutional decision-making.

Chapter 6

Discussion

Through a cross-case synthesis [Yin, 2018, pp. 194–199] of Chapters 3 through 5, I present Collaborative Narrative Scaffolding below as a basis for the design of future, similar interventions. I have framed some of my findings as design tensions with suggested design approaches, based on my work and the underlying theory. Alongside those comments, I note the places where case studies revealed but underexamined these tensions. As I return to in Chapter 7, these gaps present opportunities for future work. Following this, I discuss the limitations of my analysis and close with operational realities that may be useful for reproducing my approach in other real-world, collaborative settings.

6.1 Collaborative Narrative Scaffolding

Like Kim [2015] did with learnersourcing, I have identified certain common characteristics across these three case studies and then, in Figure 6.1, mapped these onto the foundational theory on sensemaking, scaffolding, complexity, and collaborative work that I examined in Chapter 2. In the interest of guiding future work at the intersection of these ideas, I have codified my approach into the design framework *Collaborative Narrative Scaffolding* (CNS). CNS is proposed as intermediate-level knowledge in the tradition of Höök and Löwgren; it offers a level of abstraction greater than individual instances, yet remains more

grounded than a universal theory [Höök and Löwgren, 2012]. Two of the case studies named approaches of their own. Narrative-based learnersourcing (§3.1) is the approach behind Apricot Stone City, and narrative AI infrastructure (§4.8) is the approach to system design behind CareInsights, which FITL then extended. CNS generalizes across all three, and it names what they have in common once each chapter's domain-specific terminology is set aside.

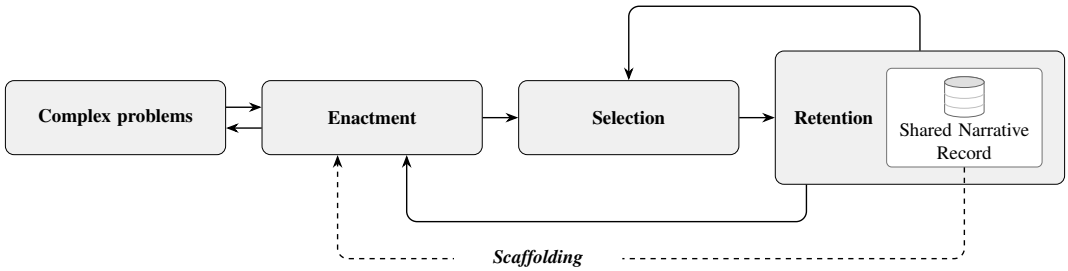
Specifically, Collaborative Narrative Scaffolding describes an applied extrapolation of Weick's Enactment-Selection-Retention (ESR) model [Weick, 1979, Weick et al., 2005]. It explicitly discusses how narrative-based data produced through sensemaking processes within a community of practice can be recorded, adapted, and deployed as personalized, context-sensitive scaffolding—leveraging a *shared narrative record*—to support members' complex problem solving. In describing CNS, I note the various ways in which AI acts as a steward of the community's shared narrative record by automating information processing burdens and curating the scaffolding. CNS assumes a community that persists long enough for one member's stories to be of use to another, and work whose sensemaking already produces those stories, so that recording them does not become a separate task. Where those conditions are absent, as in a one-off engagement or work that leaves no narrative byproduct, the shared narrative record has nothing to collect and CNS has little to offer. Because my terminology is independently derived from sensemaking, CNS could be read as a complementary design perspective with a different emphasis from how "narrative scaffolding" has been independently introduced by prior works in Chapter 2.

CNS modifies Weick's ESR model (§2.2.2) in three ways, as illustrated in Figure 6.1. Where Weick labels the first element in the diagram "ecological change" [Weick, 1979, p. 130] to describe attention-grabbing variations in experience, I instead use the framing of complex problems, as defined in §2.1.1. By my definition of the term, complex problems present a more limited-scope version of ecological change, so the original construct still holds [Weick, 1979, p. 260]. One cost of this framing is that calling something a

problem presumes it has already been noticed and acted on, which is work Weick assigns to enactment. Since my case studies only considered this particular scope, this is a more appropriate framing. The second modification is the addition of the element shared narrative record to retention as a medium of retention. This, too, is not a substantive change to the model. Rather, I am drawing attention to the importance of an externalized record of a community's stories. This falls within Weick's original framing of retention, which he notes as occurring both in individuals' minds [Weick, 1979, pp. 142–143] and in communal media [Weick, 1979, pp. 206–207]; later work traces how such externalized material carries a group from individual to group-level prospective sensemaking [Stigliani and Ravasi, 2012]. Finally, CNS highlights the role of narrative-based scaffolding. In CNS the shared narrative record is not itself the scaffolding: it persists and is revised in use [Tabak, 2004, p. 312], and what is expected to fade is the guidance drawn from it (§6.2.1). This third modification is the one I claim as new, since it adds a second path back into enactment, drawn in Figure 6.1 as scaffolding out of the shared narrative record.

Table 6.1 summarizes my cross-case synthesis to show how the specifics of my case studies map onto the framework described in Figure 6.1. To make it easier for future researchers to employ CNS, these are categorized into instantiation through infrastructuring [Le Dantec and DiSalvo, 2013, p. 247] and evaluation techniques with references that include more detailed information about how to reproduce these.

Building on existing infrastructure took its clearest form in the first case study. There, this appeared in the approach that emerged for leveraging the resources and support of TUMO, and for identifying the Yerevan cohort's surrounding community as a valuable source of stories. Apricot Stone City faced a cold start problem, since a learnersourcing platform is only as useful as the content already on it, and there was no running workflow to extend. What the community did have was its stories, so that is what the platform was built to draw on. To address this, I prompted learners to seek "stories of joy" from their own lives and from the people around them (§3.3.2), and the platform was seeded with the



Element	Description
Complex problems	Complex problems, as defined in §2.1.1, cannot be solved by routine action and often require contextual or domain-specific knowledge that is spread across different people or information systems [Dörner and Funke, 2017].
Enactment	In grappling with a complex problem, users engage in mental and interpersonal sensemaking processes. As users take action, their work changes the nature of the problems they face. Through its role in scaffolding users' sensemaking, AI has an opportunity to record interaction data that captures both what users did and their interpretive reasoning (§6.2.3).
Selection	Users are presented with opportunities to correct, inspect, and reflect on their experiences in various ways as these sensemaking byproducts are progressively archived in the form of stories. The transcription and updating of these records are designed to capture users' authentic interpretations (§6.2.2). AI assistance eases the operational and archival work of capturing, organizing, and making corrections to these narratives (§6.2.1, §6.2.3).
Retention	Within retention, as described by Weick, CNS emphasizes the use of a Shared Narrative Record. This data layer stores reusable evidence of community members' sensemaking. In addition to providing source material for scaffolding future users' sensemaking, this inspectable record has value for evaluating the system. In this regard, stories here can serve as a complement to domain-specific, paradigmatic (in Bruner's sense of the term; Bruner, 1986, pp. 11–13, 19) learning or performance outcome measures (§6.2.3).
Scaffolding	Scaffolding is delivered as narrative-based guidance that pulls from source material in the shared narrative record. The process of selecting source material and crafting personalized scaffolding is context-sensitive. This is achieved through AI management with human oversight. To avoid adding cognitive load, it is important that scaffolding is provided through the existing infrastructure (e.g., tools, activities, and social support) that users are already familiar with [Puntambekar and Kolodner, 2005] (§6.2.1).

Figure 6.1: Collaborative Narrative Scaffolding (CNS) is an adaptation of Weick's organizing model, reproduced and discussed more in Figure 2.1 [Weick, 1979, Fig. 5.1, p. 132]. As this chapter explains, this adaptation describes how scaffolding based on a shared narrative record can support enactment. The dashed arrow is the scaffolding path I add.

Case	Instantiation	Evaluation
	Complex problems	
ASC (§3.3)	Enabling young adults with little to no coding experience to effectively grasp and apply AI concepts within the decentralized structure of a learnersourcing environment [Kim, 2015]	Pre/post study AI knowledge exam; baseline self-efficacy [Pintrich et al., 1991] and readiness [Williams, 2001] surveys
CI (§4.3)	Sustaining person-centered care when caregivers are forced to navigate severe time constraints and cognitive overload without the support of usable patient records	Baseline surveys on existing tools (workload, person-centered care)
FITL (§5.3.3)	Bridging the communication and visibility gap for families who seek active involvement in a loved one's care from outside the walls of the facility	Thematic analysis of user needs by role
	Enactment	
ASC (§3.5.1)	Learners reading, authoring, and debugging stories in TUMO workshop formats, with peer help on a Discord server	Logged engagement time and level progression; UES-SF [O'Brien et al., 2018]
CI (§4.5.5.2)	Caregiver–care recipient interactions during scheduled music-therapy sessions on Memcara's tablet	Session counts and usage analytics; NASA-TLX [Hart and Staveland, 1988]
FITL (§5.4.2.4, §5.4.2.5)	Staff and family care activities within existing therapy sessions and routines, with QR anchors in resident rooms	On-site observation of workflow integration
	Selection	
ASC (§3.4.4)	Learners shaping sourced stories into scenes via workshop prompts and LLM-assisted authoring through a guidebook	Focus groups, inductively coded; per-scene ratings; expert review with a correctness and helpfulness rubric [Denny et al., 2023]
CI (§4.5.1)	Sessions recorded automatically and labeled as meaningful moments with a schema from caregivers' own notes and language	In-app caregiver ratings of generated summaries
FITL (§5.4.2.4, §5.4.2.5)	Clips drawn from existing sessions for curation, with media private by default under permissions	Think-aloud and Wizard-of-Oz prototype tests; heuristic evaluation [Ericsson and Simon, 1980, Dahlbäck et al., 1993, Nielsen, 1994]
	Retention and the shared narrative record	
ASC (§3.4.4, §3.4.5, §3.4.6)	A shared story graph seeded with prior cohorts' content, maintained through ticketing and pull-request review	Contribution logs; referral tracking; post-study community survey
CI (§4.5.3, §4.5.4)	A record built on Memcara and EHR data, with corrections attached to the data they cite	In-app feedback capture; system analytics; NPS [Alismail et al., 2020]; stakeholder interviews
FITL (§5.4.2.5)	Partner-authenticated data rendered as views modeled on cheat sheets, shift notes, and family updates	Exit interviews, including account-access friction; privacy and consent concerns tracked
	Scaffolding	
ASC (§3.2.3)	Narrative scenes and exam items teaching topics aligned to AI4K12 and AI4ALL [Touretzky et al., 2019, AI4ALL, 2023]	Pre/post isomorphic knowledge exam; UTAUT2 [Venkatesh et al., 2012]
CI (§4.5.4)	Retrieval-augmented answers modeled on peer tips [Lewis et al., 2020], with format tied to app state	In-app ratings of responses; AIM, IAM, and FIM [Weiner et al., 2017]
FITL (§5.4.2.5)	Four views fitted to staff and family roles	Exit interviews with staff and families

Table 6.1: A mapping of the three case studies into the terms of CNS. For each element of the model, this table shows how each case instantiated it through infrastructuring and the evaluation methods used, with pointers to the sections that describe them. (ASC = Apricot Stone City, CI = CareInsights, FITL = Family In The Loop)

learner-made content that the three preliminary case studies produced (§3.6.5). Here, the foundational infrastructure consisted of the stories themselves [Star and Ruhleder, 1996]. Because infrastructuring is also a proactive, forward-facing process, I posit that the shared narrative record is built intentionally for future modification by new participants [Le Dantec and DiSalvo, 2013]. This was demonstrated in ASC when the Berlin cohort adapted scenes authored by the Yerevan cohort to fix narrative gaps and create spin-offs (§3.3.3). CareInsights and FITL did not face a cold start, since both could build on routines that were already running, namely Memcara's scheduled therapy sessions (§4.5.5.2) and the cheat sheets and shift notes that staff kept themselves (§5.4.1.1).

Selection is the one element the three cases did not put in the same hands, since in Apricot Stone City learners chose what to work on and wrote it themselves (§3.4.4), while in CareInsights and FITL the system made the first pass and people worked from what it had chosen (§4.5.1, §5.4.2.5).

6.2 Design Tensions

In all three of the case studies, particular tradeoffs became apparent during the design processes that required careful attention from my collaborators and me. In the following, I summarize and direct readers to where I comment in more detail on these in Chapters 3 through 5. In isolation, these chapter discussions could be critiqued as deeply entangled with the particular intervention and the domain- and site-specific considerations of the studies. While I do not claim an overgeneralization of these findings here, through retrospective analysis, I have framed and categorized these into themes that align with the underlying terminology used in CNS. Through this approach, I hope that future researchers working on different problems and with different technology from what I have considered in this thesis may still be able to draw insights from the tradeoffs that I encountered and the design approaches that proved helpful for navigating them.

6.2.1 Context Awareness and User Privacy

There is a long tradition of HCI and CSCW research commenting on the tradeoff between awareness and user privacy [Hudson and Smith, 1996]. In all of my case studies (§4.4, §5.4.3.3), I considered this tension between building an AI intervention that intruded too much on users' privacy and intentionally not capturing data, and thus missing important contextual information that could have allowed for more personalized support—and, by extension, better outcomes and adoption. In all three cases, many of these considerations happened before the studies actually began, during the process of working with collaborators and the Dartmouth IRB to receive proper approvals for conducting human studies research. However, in the first case (working with TUMO to develop ASC) I received more clear guidance from an established, large organization with prior research experience. Once we had agreed on the terms of the study and I had received the proper approvals, TUMO provided support for communicating with families and learners and offered oversight about what data could and could not be collected and how it might be used. This facilitation, along with clear guidelines, gave users confidence to trust in the interventions and allowed the work to postpone some of the more difficult questions here that related works have noted still need answering. Thus, in Chapter 3, I underexplored this topic considering its relevance beyond my specific case.

In the other two case studies—which also did not fully resolve the tensions that emerged—this question was more front and center throughout the studies. This was not due to a lack of support from the care facilities or families, but rather, more of a reflection of the real concerns that all of the stakeholders had in this case. Making any kind of technological intervention in a long-term care facility, particularly when it involves residents living with dementia, is rife with risks and serious privacy considerations.

Design approaches:

- **Work with established partners who are experts in the field and have existing relationships with your users.** In my studies, this involved the Dartmouth IRB, TUMO (§3.3), care facilities (§4.3, §5.3.3), and Memcara (§4.5.5.2).

- **Limit the boundaries of an intervention—and in particular, data capture—to where it really matters.** A surveillance apparatus might do more harm than good. Do not let the allure of scope creep lead you there: video recording in care facilities can result in a lot of useful information that can help personalize caregiving, but it can also capture a lot more data that would be redundant or harmful (e.g., residents living with dementia removing their clothes or acting inappropriately). Once such data is recorded, an AI system might act on it in undesirable ways. To avoid such situations as best as possible in my case studies, I bounded video recording to music-based memory therapy sessions (§4.4). In my field studies, I identified these as clear opportunities for AI-assisted sensemaking scaffolding. Moreover, the meaningful moments that could be captured during these sessions, when paired with EHR data, could be more than sufficient for scaffolding support. Families and staff, who had serious concerns about continuous monitoring, were highly supportive of this approach.

- **Limit data access with role-based permissions.** This is a common suggestion [Sandhu et al., 1996], and I applied this via role-based access (§5.4.2.4, §5.4.3.3).

- **By embedding in existing infrastructure, an intervention can make its boundaries and affordances more obvious to users.** Integrate data capture into established routines and tools [Pipek and Wulf, 2009, Bossen and Markussen, 2010]; the limitations and affordances of this infrastructure can make the limitations and affordances of a new intervention more clear to users. For instance, this happened in my studies with CareInsights, where recording was embedded in Memcara's scheduled

therapy sessions, so the session itself marked where capture began and ended (§4.4, §4.5.5.2).

- **Scaffolds should fade from view when they are not needed.** While it may be valuable to continue collecting data actively, an intervention should not pester users. This happened in ASC with allowing users the flexibility to learn in various different ways and in CareInsights and FITL with how supports adapted to users' context. As prior work has framed it, guidance of this kind should fade as competence grows [van de Pol et al., 2010, Puntambekar and Hübscher, 2005], and the intervention itself should behave as what Weiser and Brown call "calm technology," staying in the periphery of users' attention and coming to the center only when needed [Weiser and Brown, 1997].

6.2.2 Particularity and Generality

As Brown and Duguid [1991, pp. 44, 47] note through their concept of *flexible generality* (§2.2.1), stories derive their instructional value in part from specific, situated details [Orr, 1996]. In my case studies, this appeared in learners' preference for stories that felt authentic (§3.6.5, §3.3.3), and in caregiving staff and family members' interest in extracting meaningful moments (§4.5.1, §5.4.2.5), which mirrored the types of notes that caregivers would themselves record if they had the capabilities to do so. However, in CNS, both the capture and the integration of stories in the shared narrative record require AI automation to be manageable at scale. The information processing that AI does in these processes risks introducing biases and generic-sounding language, as I experienced in all three case studies (§3.6.5, §4.7.3, §5.4.3). Below I summarize the different design approaches I took to navigate this tradeoff between particularity and generality.

Design approaches:

- **Stories in the shared narrative record should be a byproduct of sensemaking; they should be recorded as much as possible from the perspective of the person doing that sensemaking.** At a high level, this is the perspective that I took in all three case studies, as described in §3.3.2, §4.5.1, and §5.4.2.5.

- **Before deploying anything, work with your users to gather a data set of stories. Use this as the basis for what data in the shared narrative record should actually look like.** In ASC, this happened via how I began with storytelling workshops and then prompted learners to seek stories from their own lives and the people around them that reflected "stories of joy" (§3.3.2). For CareInsights and FITL, this involved gathering stories from user interviews and an examination of caregivers' notes. Then, I collaboratively iterated on these with end users until a clear understanding of what mattered and why could be reached.

- **Capture situated identity, and use this as a mechanism for quality control.** Situated identity is one of the properties that Weick treats sensemaking as turning on [Weick, 1995, p. 18]. Prioritize capturing stories that retain cultural specificity and human nuance (e.g., care recipients' personhood [Kitwood, 1997]), avoiding sterile, purely AI-generated summaries. Generality can be added later, when scaffolding is adapted from this material.

- **Human-in-the-loop authorship.** In all my cases, responsibility for authorship was shared, with the priority going to humans and AI filling in where users would otherwise not have the writing skills (§3.4.4), free time or cognitive capacity (§4.7.4), or firsthand access to the relevant moments (§5.4.2.5) to otherwise contribute these stories.

- **Format for the goal.** Deliver concise, "closed" summaries for immediate action (e.g., bulleted insights, §5.4.2.5), and detailed, "open" formats for reflection and

further authorship (e.g., vignettes or scenes, §3.4.4). Turner suggests that the form which persuades is not the form which coordinates [Turner, 2025, pp. 543, 546].

- **Allow new users to adapt, fix, or expand upon scaffolding based on the stories generated by other users to fit their immediate reality.** In ASC, later learners fixed plot holes and wrote spin-offs on an earlier cohort's scenes (§3.3.3).

6.2.3 Trust and Doubt

Following Weick's work on the balance of trust and doubt in organizing [Weick, 1979, p. 133], I categorize the final tradeoffs that I observed into the need for users to trust an intervention, and the importance of having users be able to doubt, inspect, and correct issues in the intervention too. Weick comments on this directly: "Has any organization ever failed to survive because it forgot something important? More likely is the possibility that organizations fail because they remember too much too long and persist too often doing too many things the way they've always done them" [Weick, 1979, p. 224]. What is true of organizations here has a corollary in AI-generated scaffolding. This is particularly true because LLMs are known to be plausible and persuasive [Bender et al., 2021, Salvi et al., 2025], which can invite overreliance [Bansal et al., 2021, Buçinca et al., 2021]. Healthy sensemaking requires doubt to be intentionally designed into the system [Reiser, 2004]. I comment on this in §3.4.4 with learner contributions reviewed before merging, in §4.7.4 with caregiver corrections to generated summaries, and in §5.4.3 with the tagging of AI-generated content.

Additionally, the relationship between what narratives plausibly reflected for the people who made them in the past and what utility they hold for a new individual's sensemaking now is a point of concern. For instance, in ASC this occurred with earlier cohorts' scenes that later learners found out of place and reworked (§3.3.3), and in CareInsights with the corrections that caregivers attached to generated summaries and the data they cited (§4.7.4). Design approaches:

- **Expose the community's interpretive process—not just final outcomes.** This provides qualitative nuance that standard outcome measures miss [Turner, 2025, p. 534] (§3.5.2.2, §4.5.1, §5.4.2.5).

- **Empower users to edit and correct underlying data.** This builds trust and permanently improves future retrieval accuracy (§4.7.4, §3.4.4).

- **Preserve authorship and traceability, even for highly synthesized content.** Cite specific sessions and notes for every AI-generated answer using retrieval-augmented generation (RAG) [Lewis et al., 2020], allowing users to audit the source (§4.5.4, §4.7.2).

- **Explicitly mark all AI-generated content.** This clearly defines the system's capabilities and limitations [Amershi et al., 2019, G1–G2] (§5.4.3).

- **Willingness to forget.** In my case studies, forgetting appeared as an organic outgrowth of user activity, interest, and editing, but this is likely not sufficient, and more work is needed here. A starting point is to treat requests to limit visibility, control access, and consent to sharing as requirements of the record rather than obstacles to it (§5.4.3.3).

6.3 Limitations

Designers and researchers interested in applying CNS should be aware of several limitations of my analysis and the evidence behind it.

To begin with, it is worth noting that this is my own reading of work I took part in rather than an outside assessment of it [Brannick and Coghlan, 2007], which departs from how a cross-case synthesis is usually positioned. I was an author on all of the case studies I examine, and my three studies were designed and published independently rather than selected in advance to replicate or contrast with one another, as the multiple-case studies

for which Yin presents cross-case synthesis normally are, though the domains themselves were chosen for maximum variation, to cover more of the design space.[1] I also arrived at CNS through retrospective analysis, so I cannot speak to how well it guides a system designed from the outset.

Secondly, there are the particular limitations of the studies themselves. I present each of these in the case studies above and would refer a reader to §3.7.6, §4.8.3, and §5.5.3 for more information on those. These included, for example, the content quality of learnersourced material, gaps in activity tracking, and a one-week evaluation window in ASC; the session-bound scope of the record and still-open questions of caregiver trust in CareInsights; and FITL's three-week deployment at four of the seven partner facilities, which could not speak to sustained engagement.

Finally, while I examined sensemaking in learning and dementia care contexts, there is a limit to how much one should expect these findings might generalize more broadly [Flyvbjerg, 2006, pp. 240–241].

6.4 Operational Realities

In all three case studies, my collaborators and I built relationships with partner organizations through sustained preliminary engagement. Once a level of trust had been established, I worked with the partner organizations and Dartmouth College to receive IRB approval. For the dementia care studies, this process took roughly half a year from the initial submission before field work could begin, and from there a further year of weekly fieldwork preceded any deployment. At the care facilities, funding cuts and administrative turnover led me to scale back studies at the affected facilities rather than pressing for their continued participation. As noted in prior HCI accounts of community-based and action research,

[1] Yin does apply the technique to cases that were conducted separately and were never planned together [Yin, 2018, pp. 194–199], and he allows elsewhere that an analytic generalization may be derived at the conclusion of a study rather than specified at its outset [Yin, 2018, pp. 37–41].

such timelines and operational hurdles are common [Le Dantec and Fox, 2015, Hayes, 2011].

I close the Discussion chapter on this not to suggest that IRB oversight should be less conservative or that deployments should proceed without first establishing trust. My case study findings, in fact, give reason for that caution, since deploying these systems surfaced unresolved concerns about privacy and consent that families and facilities are right to hold as AI inevitably begins to be integrated into care facilities (§5.4.3.3). My point is that this type of systems work—with its logistically complex deployments, evaluations, and collaborations—can provide research findings that would otherwise be out of reach. However, these realities sit awkwardly with the incentives of a PhD student, which favor publishing minimum viable papers. In part, this is why longitudinal deployment studies such as these remain rare [Long et al., 2025, Kjærup et al., 2021]. By noting these operational realities, I hope that this thesis further substantiates both the importance of longitudinal, deployment-based systems research and the case for better institutional support for such work.

Chapter 7

Conclusion

This thesis presented three deployment case studies in the domains of education and dementia care, as well as a synthesis of their generalizable insights into Collaborative Narrative Scaffolding.

7.1 Summary of Contributions

Apricot Stone City: Teaching AI Through Narrative-Based Learnersourcing

Interactive narratives have proven to be effective scaffolds for learning, and learnersourcing has emerged as a promising student-driven approach to enable personalized education and quality control at-scale. I introduced Apricot Stone City, the first synthesis of these ideas in the context of teaching artificial intelligence. Specifically, I explored the design of a narrative-based learnersourcing platform where engagement is centered around a learner-made choose-your-own-adventure story. I reported on my iterative, learner-centered design process as well as my study findings, which showed positive effects on knowledge gains, interest in AI concepts, and the overall user experience of narrative-based learnersourcing technology.

CareInsights: AI Infrastructure for Person-Centered Dementia Care

I presented CareInsights, an AI-enabled infrastructure that supports caregivers by passively capturing interaction data, enriching it with contextual and clinical signals, and generating narrative-based feedback to support day-to-day person-centered care. I evaluated CareInsights through a combination of system performance analysis, content assessments, caregiver surveys, and interviews. The caregivers who wrote the original documentation rated the generated summaries as meeting or exceeding the quality of their own, and caregivers reported outcomes consistent with more effective person-centered dementia care. Adoption was supported by integration into existing staff workflows and design choices that mirrored caregiver reasoning. I concluded with design implications for narrative AI infrastructure, offering guidance for scaling person-centered, AI-enabled support in real-world care environments. Chapter 5 widens this work's focus from professional caregivers to consider the families of people living with dementia in long-term care facilities.

Family In The Loop: Enabling Family Involvement in Dementia Care

Fragmented communication, caregiver burnout, limited resources, and regulatory complexity frequently disrupt the flow of information, creating conditions in which person-centered care becomes difficult to sustain in practice. To address these challenges, I introduced Family In The Loop, an AI-enabled web platform that enhances continuity and collaboration for person-centered dementia care. FITL automatically synthesizes diverse care data—including video, notes, and records—into contextually relevant, emotionally resonant updates tailored for staff and family members. It offers four interlinked views: Reels provides families with concise, meaningful video updates; Essentials supplies caregivers with resident-specific summaries for immediate use; Explanations presents narrative insights into care decisions; and Archives maintains a searchable, longitudinal record of care activities. Through an iterative co-creation process spanning two years and seven resource-constrained facilities, I designed and developed FITL, deploying it in real-world

dementia care contexts. Findings indicated that FITL improved staff access to critical resident information and deepened family engagement across the deployment, with caregivers describing it as no added burden. By surfacing timely, personalized, and meaningful information, FITL strengthened trust between families and staff, reduced misunderstandings, and bridged communication gaps. I concluded the chapter by discussing inherent tensions around automation, oversight, and privacy, and I proposed design considerations for adaptive, collaborative, and person-centered care technologies.

Collaborative Narrative Scaffolding

Chapter 6 brings the three case studies together and presents them as Collaborative Narrative Scaffolding (CNS), a design framework that guides the development of infrastructuring- and narrative-based AI assistance for sensemaking during complex problem solving in collaborative settings.

7.2 Future Directions

I have been glad to see AI-assisted sensemaking attract increasing attention since I began work on the case studies, most recently as the subject of a dedicated CHI 2026 workshop [Russell et al., 2026].

It would be useful for such work to further examine some of the design tensions that I commented on in Chapter 6, including the tradeoff between context awareness and user privacy (§6.2.1), the limits of summarizing stories (§6.2.2), and the risk of overreliance on AI assistance (§6.2.3). To strengthen CNS as a framework, it will be important for future work to more directly compare CNS and non-narrative AI assistance in comparable settings, establish its boundary conditions, and closely examine how AI manages the shared narrative record as community interpretations adapt over longer periods of time. The operational realities discussed in §6.4 serve as a practical call to the research community to develop

better institutional support and incentives that make longitudinal, real-world deployments more feasible to conduct.

Beyond these, it would be interesting to consider how CNS might be pushed to be more cross-institutional and more longitudinal—or, in systems thinking terms, to reconsider what is included within the bounds of the system (§2.1). To some extent, this was explored by my own case studies. I initially began Apricot Stone City (ASC) with one cohort, then progressively added others, in the process broadening the boundaries of who was a part of this "community," in Lave and Wenger's sense of a community of practice [Lave and Wenger, 1991]. Similarly, FITL was an expansion of CareInsights and expanded the system to include family members. As my case studies reported for both of these instances, complexity increases as the system size does [Roberts, 2006]. I would recommend that researchers interested in pursuing this consider how lifelong learning and connected learning might provide useful direction [UNESCO Institute for Lifelong Learning, 2022, Ito et al., 2013]. I sketched what that might look like in a workshop paper at Learning @ Scale 2025 [Moore and Murnane, 2025].

7.3 A Personal Note

I would like to end on a brief anecdote. My grandmother, Ruby, developed dementia later in life. In my experience, it made such a difference to have people who loved and deeply understood her involved in her care.

Unfortunately, as I have discussed, many individuals with dementia are right now living at resource-constrained long-term care facilities and do not receive this same level of personalized care. With further development, I am optimistic that narrative-based AI systems will be able to provide the critical sensemaking assistance that caregivers need to begin closing this gap. I encourage researchers to pursue this meaningful work.

Appendix A

Apricot Stone City: Supplementary Materials

This appendix provides supplementary materials for the Apricot Stone City study presented in Chapter 3, including study instruments, survey questions, example assessment items, and participant information.

A.1 Example Knowledge Assessment Questions

The following multiple choice questions are provided as examples of the isomorphic pre- and post-study knowledge assessments. The correct answer to each question is bolded and listed first.

- Which algorithm/model from the following list would most likely be appropriate for separating a dataset into 5 clusters?
  - **K-means**
  - Decision tree classification
  - Support vector machine

  - Naive Bayes classification
  - I don't know

- Which of the following is most suitable for classifying emails as either spam or not spam? (Assume that you already have a dataset of emails, with some that you know were marked by users as "spam".)
  - **Supervised learning**
  - Unsupervised learning
  - Reinforcement learning
  - None of the above
  - I don't know

- What is the purpose of k-fold cross-validation?
  - **To assess performance of the model on unseen data**
  - To improve the quality of training data
  - To label training data for supervised learning
  - To mix training data with test data to improve performance
  - I don't know

- Which of the following statements best describes overfitting in the context of machine learning?
  - **Overfitting occurs when a model performs well on the training data but fails to generalize to new, unseen data**
  - Overfitting is the result of a model being too simple, leading to poor performance on both training and test data

  - Overfitting is a desirable outcome, as it indicates that a model has learned the underlying patterns in the data effectively
  - Overfitting is a term used to describe the process of training a model on a diverse set of data to improve its overall performance
  - I don't know

- How might data visualizations help in identifying potential biases in training data? (Select the most relevant response)
  - **By highlighting imbalances based on demographic variables**
  - By representing data in visually appealing ways
  - By providing insights into feature correlations
  - By visualizing training time vs. model accuracy
  - I don't know

A.2 Example Learner-Made Quiz Questions

The following multiple choice questions are provided as examples of post-scene quizzes made by learners on the platform. The correct answer to each question is bolded and listed first.

- In the "practice giving a speech" example, Lia was like a reinforcement learning model and she gave a great speech—how did she learn to get better over time?
  - **By doing a smart form of trial and error and seeing what works best for her, then doing more of that thing**
  - By calculating patterns in data
  - Through the analysis of labeled training data

  - By following her teacher's instructions on how to give a great speech
  - I don't know

- In the cloud gazing scene, cloud formations represent the concept of supervised learning—why is that?
  - **Carol imagines that each cloud shape has an ideal label, like a bunny or a cat, and that with some examples she could train her friend to guess the shapes she sees**
  - Using a set of complicated logical rules in code, each cloud can be classified as a type of animal
  - When she looks at the clouds they resemble a child learning through trial and error
  - The interconnectedness of the clouds resemble a neural network
  - I have no idea

- You looked at metaphors for each neural network architecture—which one excels at processing images and extracting features?
  - **2D convolutional network**
  - Recurrent network
  - Generative adversarial network
  - Feed-forward network
  - I have no idea

- Remember Carol's interaction with the taxi driver? What part of training an AI model did her conversation and observations represent? (You'll see small hints in the scene—go back if you need help)

  - **Model evaluation**
  - Feature selection
  - Data preprocessing
  - Model training
  - Idk, take me back to the scene!

A.3 Pre- and Post-Study Survey Instruments

This section provides questions from the pre- and post-study surveys. Response options are bold, with information about the metric in italics.

A.3.1 Questions Asked Pre- and Post-Study

- Academic engagement and self-efficacy

  Select how much you agree or disagree with the following statements. **[Strongly disagree, Disagree, Neutral, Agree, Strongly agree]** *(Likert scale, range: [1, 5])*

  - I am motivated to learn about AI.
  - If I wanted to, I could potentially do very well in computer science.
  - I try to make connections between what I learn in different classes and experiences.
  - I put a lot of effort into the work I do.
  - Even when things are tough, I can perform quite well.

- Self-efficacy

  Select how much you agree or disagree with the following statements. **[Strongly disagree, Disagree, Neutral, Agree, Strongly agree]** *(Likert scale, range: [1, 5])*

  - I'm confident I can understand the basic AI concepts taught in this study.

  - I'm confident I can understand the most complex AI material presented in this study.
  - I believe I can do an excellent job on the AI-related activities and evaluations in this study.
  - I'm certain I can master the AI skills being taught in this study.

- Task-value

  Select how much you agree or disagree with the following statements. **[Strongly disagree, Disagree, Neutral, Agree, Strongly agree]** *(Likert scale, range: [1, 5])*

  - I think I will be able to use what I learn in this study in other classes.
  - It is important for me to learn the material in this study.
  - I am very interested in the topics of this study.
  - I think the material in this study is useful for me to learn.
  - I like the subject matter of this study.
  - Understanding the subject matter of this study is very important to me.

- Interest value, utility value, attainment value

  Select how much you agree or disagree with the following statements. **[Strongly disagree, Disagree, Neutral, Agree, Strongly agree]** *(Likert scale, range: [1, 5])*

  - I like AI.
  - AI is exciting to me.
  - I am fascinated by AI.
  - AI concepts are valuable to learn.
  - Being good at AI will be important when I get a job or go to college.
  - Being someone who is good at computer science is important to me.

- Belongingness in computer science, sense of social academic fit

  Select how much you agree or disagree with the following statements. **[Strongly disagree, Disagree, Neutral, Agree, Strongly agree]** *(Likert scale, range: [1, 5])*

  - I feel a sense of belonging to the AI and computer science community.
  - I feel comfortable in computer science.
  - I feel like an outsider in computer science.
  - I identify as a computer scientist.

- Readiness to learn

  Select how much you agree or disagree with the following statements. **[Strongly disagree, Disagree, Neutral, Agree, Strongly agree]** *(Likert scale, range: [1, 5])*

  - I am good at setting goals and deadlines for myself.
  - I finish things I start.
  - I do not quit just because things get difficult.
  - I am relatively good at using the computer.

- Storytelling enjoyment

  Select the response you most agree with. **[Not at all, A little, Quite a bit, Very much]** *(Likert scale, range: [1, 4])*

  - How much do you enjoy reading stories in general?
  - How much would you enjoy reading stories as part of the study?
  - How much do you enjoy writing stories in general?
  - How much would you enjoy writing stories as part of the study?
  - What is your favorite book? **[Free response]**

- Prior experiences with computer science, AI, and storytelling

  Select the response you most agree with. **[Yes, No]**

  - I have participated in a computing extracurricular activity before.
  - I have taken a computer programming or AI class before.
  - I have participated in a storytelling-related extracurricular activity before.
  - I have taken a class before that involved reading and/or writing stories.

- Self-reported knowledge of AI topics covered on the platform

  Reviewing the following statements, rate your familiarity with the concepts discussed. **[I have written code that relates to this, I could explain this concept to a friend, I am somewhat familiar with this concept, I could guess what this means, I don't know what this means]** *(Likert scale, range: [1, 5])*

  - Define supervised, unsupervised, and reinforcement learning algorithms, and give examples of human learning that are similar to each algorithm.
  - Model how machine learning constructs a reasoner for classification or prediction by adjusting the reasoner's parameters (its internal representations).
  - Use either a supervised or unsupervised learning algorithm to train a model on real-world data, then evaluate the results.
  - Illustrate what happens during each of the steps required when using machine learning to construct a classifier or predictor.
  - Describe how various types of machine learning algorithms learn by adjusting their internal representations.
  - Select the appropriate type of machine learning algorithm (supervised, unsupervised, or reinforcement learning) to solve a reasoning problem.

  - Train a multi-layer neural network using the backpropagation learning algorithm and describe how the weights of the neurons and the outputs of the hidden units change as a result of learning.
  - Compare two real-world datasets in terms of the features they comprise and how those features are encoded.
  - Evaluate a dataset used to train a real AI system by considering the size of the dataset, the way that the data were acquired and labeled, the storage required, and the estimated time to produce the dataset.
  - Investigate imbalances in training data in terms of gender, age, ethnicity, or other demographic variables that could result in a biased model, by using a data visualization tool.
  - How to use variables in a programming language.
  - What JavaScript is.
  - The difference between TypeScript and JavaScript.
  - Using conditionals, like “if”, in code.
  - Using loops, like “while” and “for”, in code.
  - Using a map/dictionary data structure in code.
  - Representing a graph using a data structure in code.
  - Using git to push and pull commits.
  - Using git to resolve a merge conflict.
  - Using Node.js.

- Peer comparison
  - Compared to your peers, how would you rate your knowledge of computer science? **[Much less knowledgeable, Somewhat less knowledgeable, Average,**

Somewhat more knowledgeable, Very knowledgeable] *(Likert scale, range: [1, 5])*

- Demographics
  - How many years old are you? **[Number drop down]**
  - What is your gender? **[Female, Male, Non-binary, Free response]**
  - What is your nationality? **[Free response]**
  - What is your racial/ethnic identity? **[Free response]**
  - Do you identify as Armenian? **[No, Yes, Free response]**
  - Who (if anyone) referred you to this opportunity? **[Free response]**

A.3.2 Questions Asked Only Post-Study

- Reflections on the study
  - How important was it to you that the story focused on Armenia? Please explain. **[Free response]**
  - Did it matter that the story was/wasn't relevant to your personal experiences? Please explain. **[Free response]**
  - What originally motivated you to sign up for this study? What kept you motivated to keep participating? Please elaborate. **[Free response]**
  - I would choose to participate in an AI learning community again. **[Yes, No]**
  - I feel that I am a member of the Apricot Stone City learning community. **[Yes, No]**
  - The Apricot Stone City learning community is supportive of me. **[Yes, No]**
  - Did you feel like there was a sense of community during the study? **[Yes, No]** Please explain. **[Free response]**

- User Engagement Scale Short Form [O'Brien et al., 2018], scoped to questions on focused attention and reward factors. **[Strongly disagree, Disagree, Neutral, Agree, Strongly agree]** *(Likert scale, range: [1, 5])*
  - I was absorbed in this experience.
  - My experience was rewarding.
  - I felt interested in this experience.
- Extended Unified Theory of Acceptance and Use of Technology [Venkatesh et al., 2012], scoped to questions on performance expectancy, effort expectancy, social influence, facilitating conditions, hedonic motivation, habit, and behavioral intention. **[Strongly disagree, Disagree, Neutral, Agree, Strongly agree]** *(Likert scale, range: [1, 5])*
  - This experience helped me learn.
  - Learning how to use the Apricot Stone City platform was easy for me.
  - People who are important to me would want me to engage with the platform.
  - I have the resources necessary to use the platform.
  - Using the platform was enjoyable.
  - Using the platform would become a habit for me.
  - I would intend to continue using the platform in the future if it's available.
- Self-reported learning per level

  Please rate the following activities in terms of how much you feel you learned by engaging with them. (If you didn't engage with a level, mark N/A). **[I didn't find this at all valuable for learning, It was a little valuable for learning, It was very valuable for learning, N/A (I didn't engage with this)]** *(Likert scale, range: [1, 3] or N/A value)*

  - Level 1 (using existing content on the platform)
  - Level 2 (creating content)
  - Level 3 (coding mini-games)
  - Level 4 (bringing the concepts into the real-world, considering AI ethics)
  - After level 4 (supporting other content creators)
  - How far did you progress on the site (e.g., what level)? What helped you get that far? What prevented you from getting farther? Please explain. **[Free response]**

A.4 Post-Scene Feedback Questions

A learner received the following questions to provide feedback after completing a story scene.

- Enjoyment and self-reported learning
  - How enjoyable was this scene? **[Animated smiley emoticon Likert scale]**[1]
  - How would you rate your learning for this scene? **[Animated smiley emoticon Likert scale]**
  - What did you like about this scene? **[Free response]**
  - What would you change about the scene? **[Free response]**

A.5 Focus Group Interview Guide

The semi-structured focus group interviews used the following questions as a guide, centered around the research questions.

- Questions related to RQ1

[1] `https://www.npmjs.com/package/react-rating-emoji`

  - Tell me about a time when you felt like you were learning something new and interesting on the platform.
  - What were some of the favorite new things you learned this past week? It could be AI concepts, or how to use some aspect of GitHub—or anything else—whatever you enjoyed learning!
  - Now, how about a time you felt confused or found something hard to understand?
  - How did this experience compare to regular school and learning in a classroom with a teacher?

- Questions related to RQ2
  - How enthusiastic were you about AI / computer science before this experience? Has that interest gone up or down?
  - Do you think you'll take more AI or computer science classes in school or after-school programs if you have the chance? Why or why not?

- Questions related to RQ3
  - Tell me about some of the things you liked the most about the platform/your experience?
  - Anything you really disliked or would want to change?
  - How did you feel about the stories on Apricot Stone City?
  - What was your favorite scene? Why?
  - Were there any stories you really disliked? How come?
  - Did you get to the point of writing story content? How was that experience?
  - Did you feel like you were part of a community on Apricot Stone City? Tell me about that.

  - How much did you interact with other learners on the platform?
  - What encouraged you or discouraged you from interacting with others?
  - Did you feel ownership of the platform content?
  - How did the bugs in the platform content make you feel? (e.g., empowered to fix them? / did you notice them?)

- Questions related to scalability
  - If you had one more month to continue working on this platform, what would you like to see happen?
  - Do you imagine you'll keep using the platform once the official study ends? Why or why not? It's OK if the answer is no—we want to understand your honest reactions!
  - Would you refer the platform to a friend?
  - Could you see the platform being used in a classroom context? At home?

A.6 Participant Information

Tables A.1 and A.2 provide basic demographic information (age and gender) about all study participants. The "Group" column indicates participation in one of the preliminary case studies (C1, C2, C3) or the final system evaluation (Study).

Table A.1: Participant demographics for the three preliminary case studies (C1, C2, C3).

ID	Age	Gen.	Grp.
C1-1	13	F	C1
C1-2	14	F	C1
C1-3	15	F	C1
C1-4	15	F	C1
C1-5	16	F	C1
C1-6	16	F	C1
C1-7	16	F	C1
C1-8	16	F	C1
C1-9	17	F	C1
C1-10	12	M	C1
C1-11	14	M	C1
C1-12	14	M	C1
C1-13	14	M	C1
C1-14	16	M	C1
C1-15	16	M	C1
C1-16	16	M	C1
C1-17	17	M	C1
C1-18	17	M	C1
C2-1	16	F	C2
C2-2	16	F	C2
C2-3	16	F	C2
C2-4	16	F	C2
C2-5	16	F	C2
C2-6	16	F	C2
C2-7	16	F	C2
C2-8	17	F	C2
C2-9	17	F	C2
C2-10	17	F	C2
C2-11	17	F	C2
C2-12	17	F	C2
C2-13	17	F	C2
C2-14	17	F	C2
C2-15	17	F	C2
C2-16	17	F	C2
C2-17	17	F	C2
C2-18	17	F	C2
C2-19	18	F	C2
C2-20	18	F	C2
C2-21	19	F	C2
C2-22	15	M	C2
C2-23	15	M	C2
C2-24	15	M	C2
C2-25	16	M	C2
C2-26	16	M	C2
C2-27	16	M	C2
C2-28	16	M	C2
C2-29	16	M	C2
C2-30	17	M	C2
C2-31	18	M	C2
C3-1	15	F	C3
C3-2	16	F	C3
C3-3	16	F	C3
C3-4	16	F	C3
C3-5	16	F	C3
C3-6	16	F	C3
C3-7	17	F	C3
C3-8	17	F	C3
C3-9	14	M	C3
C3-10	15	M	C3
C3-11	15	M	C3
C3-12	16	M	C3
C3-13	16	M	C3
C3-14	16	M	C3
C3-15	17	M	C3
C3-16	17	M	C3

Table A.2: Participant demographics for the final system evaluation study.

Participant ID	**Age**	**Gender**	**Group**
P1	14	F	Study
P2	17	F	Study
P3	17	F	Study
P4	17	F	Study
P5	17	F	Study
P6	18	F	Study
P7	18	F	Study
P8	19	F	Study
P9	19	F	Study
P10	20	F	Study
P11	20	F	Study
P12	20	F	Study
P13	23	F	Study
P14	16	M	Study
P15	17	M	Study
P16	18	M	Study
P17	18	M	Study
P18	18	M	Study
P19	18	M	Study
P20	18	M	Study
P21	19	M	Study
P22	19	M	Study
P23	19	M	Study
P24	21	M	Study
P25	21	M	Study
P26	22	M	Study
P27	23	M	Study

Appendix B

CareInsights: Supplementary Materials

This appendix provides supplementary materials for the CareInsights study presented in Chapter 4, including caregiver-generated vignettes from the formative annotation process, expanded technical descriptions of the Capture and Synthesis APIs, example system outputs, person-centered care survey items, and care recipient demographics.

B.1 Caregiver-generated Vignettes from Thematic Video Analysis

This section describes how caregivers made their tacit reasoning explicit and, in doing so, helped formalize a mental model for tracking engagement and recognizing Meaningful Moments. The aim was to surface the cues caregivers attend to, how they weigh those cues for a given resident, and how those judgments accumulate into short summaries of clinically relevant events.

Caregivers first rewatched a recorded music-therapy session and wrote brief retrospective care notes that identified candidate Meaningful Moments. They then replayed the video and annotated a simple timeline with the engagement dimensions they judged most diagnostic, explaining aloud why each visible change mattered for the resident's state and

care needs. The set of dimensions was intentionally open ended, but caregivers consistently converged on five families of markers: attention, emotional expression, memory recall, reactions to activities, and behavioral symptoms. Through discussion, they also calibrated resident-specific thresholds and clarified how conversational content, observable behavior, and biographical context interact to signal a meaningful change.

These explanations and timelines yielded a practical, caregiver-grounded annotation schema that captured both what to mark and why. Using this schema, caregivers labeled 203 videos; those labels were then used to train the model (see Section B.2). As reported in Section 4.7, the resulting system reflects the reasoning patterns articulated in this process and reproduces caregiver judgments about Meaningful Moments.

Two pseudonymized caregiver-authored timeline vignettes are included below, drawn directly from caregivers' own work during the annotation process. These examples, Adam (F2-CR11, mild impairment) and Birdie (F1-CR6, advanced dementia), illustrate how caregivers interpreted engagement cues and organized their reasoning when identifying and describing Meaningful Moments across different cognitive profiles.

B.1.1 Adam: Verbal Cues of Memory Recall

In this music-therapy session with Adam, a verbal resident with mild cognitive impairment, the caregiver identified two segments showing elevated memory recall based entirely on the conversation transcript. The first occurred when a song with a distinct snare drum rhythm prompted Adam to reveal that he used to be in a band. This was something the caregiver had never been briefed on, despite its clear relevance to music-based therapy.

Adam: I'm a drummer.

Therapist: Is that right?

Adam: Yeah.

The second Meaningful Moment came later, when Adam began talking about a former bandmate who played the trumpet. Being able to detect and share these events passively is valuable for care teams, as details like these help personalize therapy and strengthen continuity across sessions.

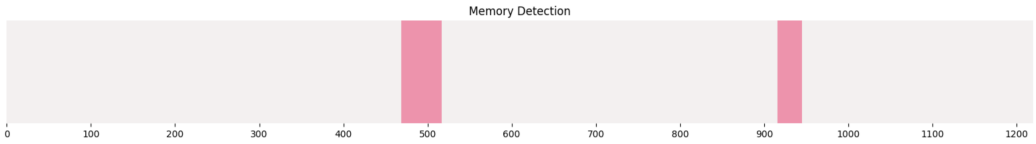


Figure B.1: Excerpt of the caregiver-generated timeline for Adam's session, showing expert annotations of the "Memory Recall" dimension.

B.1.2 Birdie: Cognitive and Emotional Engagement

Birdie's case shows how Meaningful Moments can occur even when a memory is not being directly shared by the care recipient, as in Adam's case. In this vignette, the indicators reflect cognitive engagement and enjoyment. During her first two music-therapy sessions (timelines not shown), Birdie showed little response. In the caregiver's notes from these sessions, they remarked that music therapy seemed ineffective for her and questioned whether it was worth continuing. By the third session (Figure B.2), however, certain songs began to evoke more consistent engagement. When the therapist played *Fly Me to the Moon* by Frank Sinatra and began singing along, Birdie started tapping her hand in rhythm and soon joined in, singing parts of the song. Segments of the session when she was visibly smiling are marked on the "Emotion Pleasant" timeline, and segments when she began asking questions or talking with the caregiver are marked on the "Reaction Talking" timeline.

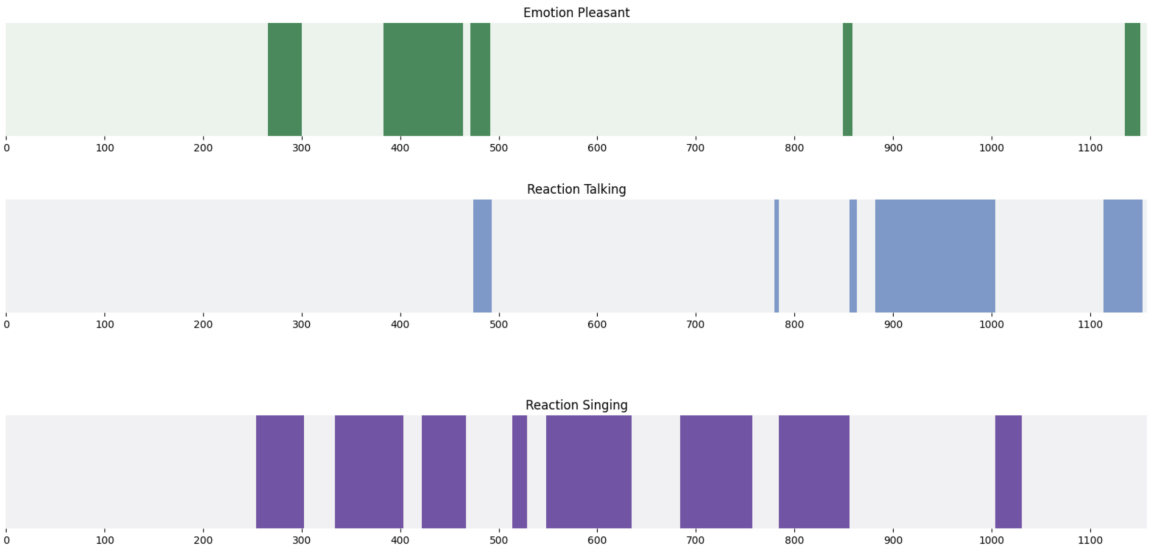


Figure B.2: Excerpt of the caregiver-generated timeline for Birdie's session, showing expert annotations of "Emotion Pleasant," "Reaction Talking," and "Reaction Singing."

As the session continued, Birdie began talking more with the therapist. At 880 and 1118 seconds, she said, "I love you" and later, "I liked singing with you," which the caregiver tagged as particularly notable for Birdie, given that she was generally withdrawn. The pattern captured by the sequence of Meaningful Moments here, and the absence of them in earlier sessions, demonstrates that even residents who initially show little response can become expressive when therapy is sustained and emotionally meaningful.

B.2 Expanded Capture API Technical Description

High-quality documentation is essential for supporting person-centered dementia care, but existing approaches often impose unsustainable burdens on caregiving staff. CareInsights addresses this challenge by embedding lightweight, passive data capture directly into existing care routines, minimizing additional workload while enriching the available information about residents. The approach integrates passive video recording during therapy sessions with automatic annotation of engagement-related behaviors, creating a rich observational layer without requiring caregivers to manually document every interaction.

Passive video capture through platform integration and the lightweight gold-standard labeling system are described in Chapter 4. This section expands on the remaining two parts of the capture infrastructure: the multimodal vision-language analysis pipeline, and change-point detection and summarization for Meaningful Moment extraction.

Motivation for a Stacked Ensemble Approach Caregiving interactions involve a rich interplay of visual cues, conversational content, emotional expression, and behavioral dynamics. No single modality consistently captures all relevant engagement signals. To address this challenge, I adopted a stacked ensemble design [Kunapuli, 2023], where specialized base learners process different aspects of the session data, and a meta-learner combines their outputs into final predictions. This structure leverages the complementary strengths of individual modalities while mitigating weaknesses or biases that any one model might introduce.

Feature Extraction Across Modalities Each session recording is automatically decomposed into aligned video and audio streams. These are processed independently to extract modality-specific features, ensuring that each base learner focuses on a consistent temporal granularity of 30-second windows, except for the transcript stream, which uses overlapping 60-second windows to preserve conversational context.

Visual Features. Video frames are sampled at 30 frames per second and processed using two pretrained computer vision models. The `deepface` library [Serengil and Ozpinar, 2021] extracts emotion probabilities across seven categories—anger, disgust, fear, happiness, sadness, surprise, and neutrality—on a per-frame basis. While I observed systematic misclassifications when applying DeepFace to dementia care footage (e.g., muted engagement behaviors misread as negative emotions), I found that these biases were consistent and thus useful as features for a downstream meta-learner. Emotion probabilities are aggregated across each window by computing both mean and variance values, capturing average affective state as well as emotional dynamism. In parallel, MediaPipe [Lugaresi et al., 2019]

extracts facial landmarks to compute eye closure ratios, serving as a proxy for drowsiness and attentional shifts. These features are similarly averaged and variance-calculated across each window.

Audio and Transcript Features. The audio stream is transcribed using Whisper [Radford et al., 2023], and dialogue is attributed to specific participants through a separate diarization step. The transcript is segmented into overlapping windows of approximately 60 seconds, chosen to capture coherent stretches of dialogue while maintaining alignment with behavioral patterns detected in the video stream. This windowing approach ensures that each segment contains enough conversational context for meaningful analysis without drifting too far from the associated visual cues.

To extract higher-level semantic information from each window, GPT-4o is applied using a structured prompt developed through iterative testing with caregivers. The goal is to generate interpretable features that reflect the resident's cognitive, emotional, and behavioral engagement during the session. Each prompt guides the model to return quantitative ratings across several predefined dimensions that mirror the indicators surfaced during the earlier manual review. The following is an abbreviated version of the specific prompt.

You are assisting in analyzing transcript segments from music-based memory therapy sessions involving older adults with memory loss. Your task is to review the segment and assign a numerical rating (1–100) for each of the engagement indicators below. A score of 100 means the behavior is clearly present; a score of 1 means it is clearly absent.
Use the following information:

- **Transcript segment:** [INSERT TRANSCRIPT HERE]
- **EHR summary:** [INSERT EHR SUMMARY HERE]
- **Stimuli shown during this window:** [INSERT STIMULUS LOGS]

Rate each of the following, preserving the order listed:

- Likelihood that the resident is recalling a personal memory
- Likelihood that the resident appears emotionally sad
- Likelihood that the resident appears emotionally happy
- ...
- Likelihood that the resident shows signs of confusion
- Likelihood that the resident shows signs of behavioral disturbance
- Likelihood that the resident shows signs of withdrawal
- ...
- Likelihood that the resident is positively responding to a conversation prompt
- Likelihood that the resident is negatively responding to a conversation prompt
- Likelihood that the resident is positively responding to a song
- ...

Output format: Return ONLY a single comma-separated list of numeric scores in the same order as above. Do not include any labels, explanations, or extra text.

This method produces a structured vector of semantically meaningful features for each transcript window. These values complement the visual feature streams by capturing affective tone, memory access, conversational responsiveness, and behavioral cues that are often conveyed verbally. By framing these outputs as continuous probabilities, the system enables the meta-learner to interpret nuanced engagement dynamics beyond rigid classifi-

cation. To avoid redundant computation, outputs are cached for each window and reused if the prompt remains unchanged.

Meta-learner Training and Prediction The outputs from the visual and language base learners are concatenated to form a feature vector for each 30-second window. An XGBoost classifier is trained as the meta-learner, using caregiver-provided labels as supervision. The model predicts second-by-second engagement estimates for each of the five factors: attention, emotional valence, memory recall, reactions to therapy activities, and dementia-related symptoms. The meta-learner's design allows it to flexibly weight different modalities depending on context. For example, in highly verbal sessions, transcript features may dominate, while in quieter sessions, visual attention cues may be more informative. Multiple structured ensemble classifiers were tested to identify a model capable of handling diverse features from text and video. Comparing performance, XGBoost was selected, as it slightly outperformed alternatives like Random Forest and offered interpretability suited for the mixed-modality dataset.

B.2.1 Identifying and Describing Meaningful Moments

The Capture API first detects shifts in continuous engagement signals, marking the associated segments as candidate events. It then applies heuristics based on caregiver-authored descriptions to classify each segment as a Meaningful Moment. Finally, it generates a description for each Meaningful Moment.

Detecting Candidate Moments via Change Points Potential event boundaries are identified by applying change point detection to per-second engagement predictions. Each engagement factor is analyzed using a least-squares deviation cost function to locate shifts that indicate a change in resident state. These inflection points form candidate anchors for Meaningful Moments. Next, each segment is classified using heuristics derived from

caregiver-authored descriptions introduced earlier. For instance, a sustained rise in the memory recall signal may correspond to a moment of reminiscence, whereas a spike in behavioral symptoms followed by a decrease may reflect successful redirection. These heuristics were informed by patterns observed across manually annotated sessions and refined through earlier prototype development.

Generating Meaningful Moment Descriptions Once labeled, each segment is passed through a GPT-4o-based synthesis step. For each event type, a tailored prompt template is populated with contextual information: the segment's transcript, engagement factor values, and relevant stimuli or metadata (e.g., songs played or caregiver prompts used). These prompts guide the model to produce 2–3 sentence summaries in natural language that match the tone and intent of caregiver documentation. The structured templates were adapted from those validated in feasibility testing, where caregivers found prompt-guided summaries to be as useful and accurate as their own notes. Encoding that logic into a rule-based framework allowed the process to scale while maintaining alignment with caregiver expectations.

Structured Output for Downstream Use The Capture API produces structured annotation data that links raw video input to temporally aligned interpretations of engagement and behavior. For each session, the system outputs:

- A transcript of the video
- Per-second predictions for each of the five engagement factors
- Detected change points and labeled moment types
- GPT-4o–generated narrative summaries for each Meaningful Moment

These outputs are time-aligned, enabling downstream systems or human reviewers to trace how behavioral patterns evolve throughout a session and how specific segments were

interpreted. This layered representation provides a clear, interpretable bridge between unstructured video and structured care-relevant insight.

B.3 Expanded Synthesis API Technical Description

B.3.1 Retrieving and Filtering Relevant Context

The Synthesis API draws from a shared data layer that aggregates narrative summaries from the Context API, engagement features and Meaningful Moment descriptions from the Capture API, and prior caregiver queries and feedback. This content is filtered using batched GPT-4o prompts to identify segments relevant to the current query. Relevance is ranked using metadata such as timestamps, data type, and linked caregiver reviews, ensuring that responses are grounded in timely and meaningful context.

B.3.2 Prompt Construction and GPT-based Synthesis

The filtered content is then injected into a structured GPT-4o prompt. This prompt includes the original caregiver query, followed by a block of retrieved context snippets, each labeled with a reference ID. The model is instructed to cite these sources explicitly using inline LaTeX-style syntax and to favor explanations that are directly supported by the input data.

To support downstream parsing, the prompt structure remains consistent across queries. Citations to video content specifically refer to previously detected and labeled Meaningful Moments rather than entire video sessions. Because Meaningful Moments are timestamped, references directly link to specific segments within longer video recordings, providing the option for clinicians to efficiently verify underlying observations, revisit critical interactions, and surface illustrative examples for reflection or training.

B.3.3 Post-processing and Citation Validation

After the response is generated, it is parsed into discrete text segments and citation tags. Each citation is then validated by checking whether the associated statement is supported by a specific segment from a referenced source. While this approach involves GPT-4o reviewing its own outputs (a setup with known limitations), it proved effective in this scenario, where the validation task is narrowly scoped and grounded in targeted comparisons of text segments. The primary challenge encountered was formatting inconsistencies—cases where generated outputs deviated from the expected structure, complicating parsing. These were resolved through iterative prompt tuning and post-processing.

Responses that lack citations are intentionally retained, which typically occur when limited data is available on a resident. Although these explanations tend to be more general, caregivers often still found them useful for orienting to a situation or confirming best practices. While outside the scope of this work, the presence or absence of citations could serve as a useful signal in future versions of the system to distinguish between personalized insights and broader, evidence-informed guidance.

B.4 Example Outputs from the Synthesis API

The following examples illustrate Synthesis API responses to caregiver queries. In these examples, questions were issued just before a 1:1 session began, with the format parameter set to produce concise outputs emphasizing key details and prioritizing actionable guidance. Redactions have been applied to preserve anonymity, and citations have been abbreviated for readability.[1]

Prompt:

[1] Superscripts (e.g., [n]) map to items in the cited block (see "Cited:" below); each citation lists the source type (e.g., EHR, Meaningful Moment) with a tap-through link in the user interface to the underlying source data.

I've never worked with this person[2] before, what should I know? I'm thinking of playing music to calm her down, any suggestions? Should I mention any family members?

Response: <*Redacted Name*> was born in <*Redacted Year*>[1]. She speaks English and lives in room <*Redacted Number*>[1]. Her MoCA score is normal, she's highly communicative, and has good hearing and eyesight[1].

When engaging with her, you can talk about topics like Hungary, her family, Broadway, and delicious Hungarian food[1, 2, 3, 4]. She also enjoys shopping and discussing local restaurants[5]. Discussing her grandchildren can be a great way to connect[1, 3]. Incorporating music into sessions can be immensely soothing and enjoyable for her[6]. She loves pop music[1, 6, 7, 8, 9]. Songs like *"I Have a Dream"* by ABBA and *"Stardust"* by Nat King Cole can evoke positive memories[7]. Avoid unfamiliar artists, and focus on joyful music that she recognizes[10].

She enjoys cooking, gardening, and games[1, 3, 11]. In terms of redirection, music or being with people helps refocus her[6, 7]. If she's ever feeling distressed, repetitive affirmations like saying "okay" can provide reassurance[12]. Avoid serving her poorly prepared or bad-tasting food as this could be upsetting for her[16].

Cited: **[1]** EHR short summary: <*Redacted Name*> was born in <*Redacted Year*>. She speaks English and Hungarian. She lives in room number <*Redacted Number*>. Hearing is normal levels for both ears. Eyesight is normal. Medical history: <*Redacted Text*> Current medications: <*Redacted Text*> Family info: <*Redacted Text*> ...

[...]

[6] Meaningful Moment: <*Video Clip ID*>:<*Timestamp*> Memory recall of <*Redacted Name*>, brother, after listening to ABBA's "*I have a dream*" during music therapy program.

[2]CareInsights identifies the referenced user through the Memcara platform's care recipient selection feature, which indexes into the appropriate context window. Asking a query is only possible once a user is selected.

[7] *<Video Clip ID>:<Timestamp>* Heightened positive engagement when caregiver *<Redacted Therapist Name>* discusses Nat King Cole's hit song "*Stardust*" with *<Redacted Care Recipient Name>*.

[...]

Prompt:

> *<Redacted Name>* is agitated, I am the only one around and I need to get her to eat some food. Help. I really need calm.

Response: When *<Redacted Name>* is agitated, try talking to her about her children and family[1, 2, 3]. This topic often brings her comfort and helps redirect her focus[4]. Discussing animals and pets may also calm her as she has shown interest in these topics[5, 6]. Avoid any discussion of her ex-husband as it could increase her agitation[7].

Instead of using music, which can sometimes lead to overstimulation if not managed carefully, gently read aloud to her as this can be soothing and may help her relax[8, 9, 10, 11]. Avoid any environments with lots of noise to prevent exacerbating her distress[12]. If you feel it is safe and suitable, you can talk about songs the resident enjoys, like those by Kay Starr, which may prompt positive memories without needing to play the music[12, 13, 14, 15].

Cited: [...]

B.5 Survey Items for Person-centered Care

These items were adapted from the Person-centered Care Assessment Tool (P-CAT) [Edvardsson et al., 2010] to reflect the role of digital documentation tools in facilitating personalized care strategies for residents with dementia. Based on caregiver feedback, the availability of life story information was emphasized, as well as how accessible, useful,

and actionable that information is during daily care tasks. All responses were recorded on a standard 7-point Likert scale (1 = Strongly disagree, 7 = Strongly agree).

1. I believe that understanding a person with dementia's life stories, topics of interest, and personal preferences is essential for providing effective care.
2. The care documentation tool captures comprehensive information about residents' life stories, interests, and preferences.
3. I utilize the care documentation tool to record and access information about a care recipient's personal history and preferences.
4. The care documentation tool provides an appropriate amount of information about care recipients' personal backgrounds and preferences.
5. It is easy to search for detailed information about a care recipient's life history and preferences using the care documentation tool.
6. The care documentation tool integrates information about care recipients' personal histories into care strategies, such as approaches to redirecting challenging behaviors.
7. The care documentation tool highlights the most critical information about a care recipient's life story and preferences.

B.6 Care Recipient Demographics

Table B.1 provides details of care recipients at the facilities included in all components of the final analysis. Three other facilities experienced unexpected administrative departures and funding reductions; at those sites deployment support continued but other research activities were scaled back during evaluation.

Table B.1: Details of care recipients at facilities (F) included in all components of the final analysis.

Participant ID	Facility ID	Progression of dementia	Age	Gender
F1-CR1	F1	None	75	F
F1-CR2	F1	Moderate	89	F
F1-CR3	F1	Moderate	91	F
F1-CR4	F1	Advanced	85	F
F1-CR5	F1	Advanced	86	F
F1-CR6	F1	Advanced	89	F
F1-CR7	F1	Advanced	90	F
F1-CR8	F1	Advanced	91	F
F1-CR9	F1	Advanced	92	F
F2-CR1	F2	None	79	F
F2-CR2	F2	None	62	F
F2-CR3	F2	None	79	M
F2-CR4	F2	None	88	F
F2-CR5	F2	None	95	F
F2-CR6	F2	Mild	75	F
F2-CR7	F2	Mild	76	F
F2-CR8	F2	Mild	83	F
F2-CR9	F2	Mild	80	F
F2-CR10	F2	Mild	76	F
F2-CR11	F2	Mild	84	M
F2-CR12	F2	Mild	79	F
F3-CR1	F3	Mild	58	F
F3-CR2	F3	Mild	75	F
F3-CR3	F3	Moderate	75	F
F3-CR4	F3	Moderate	84	F
F3-CR5	F3	Moderate	85	F
F3-CR6	F3	Moderate	94	F
F4-CR1	F4	Mild	70	F
F4-CR2	F4	Mild	80	M
F4-CR3	F4	Moderate	70	F
F4-CR4	F4	Advanced	77	F
F4-CR5	F4	Advanced	81	F
F4-CR6	F4	Advanced	90	F

Appendix C

Family In The Loop: Supplementary Materials

C.1 Prototype Gallery

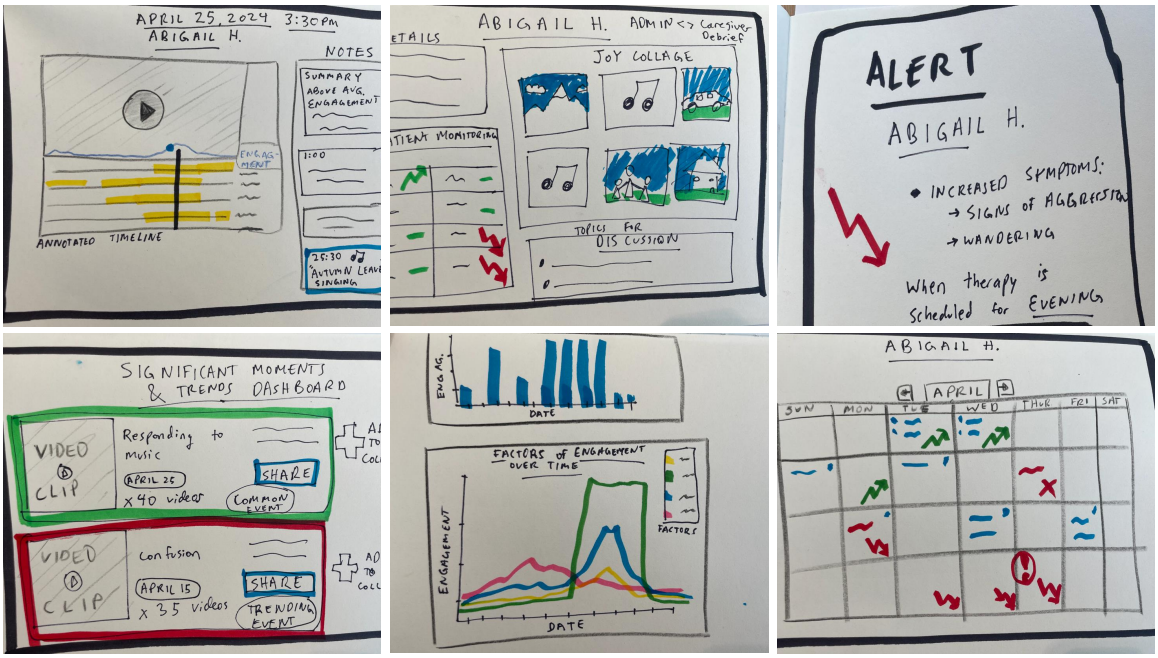


Figure C.1: Excerpts from early low-fidelity prototypes. All user content shown here is mocked data.

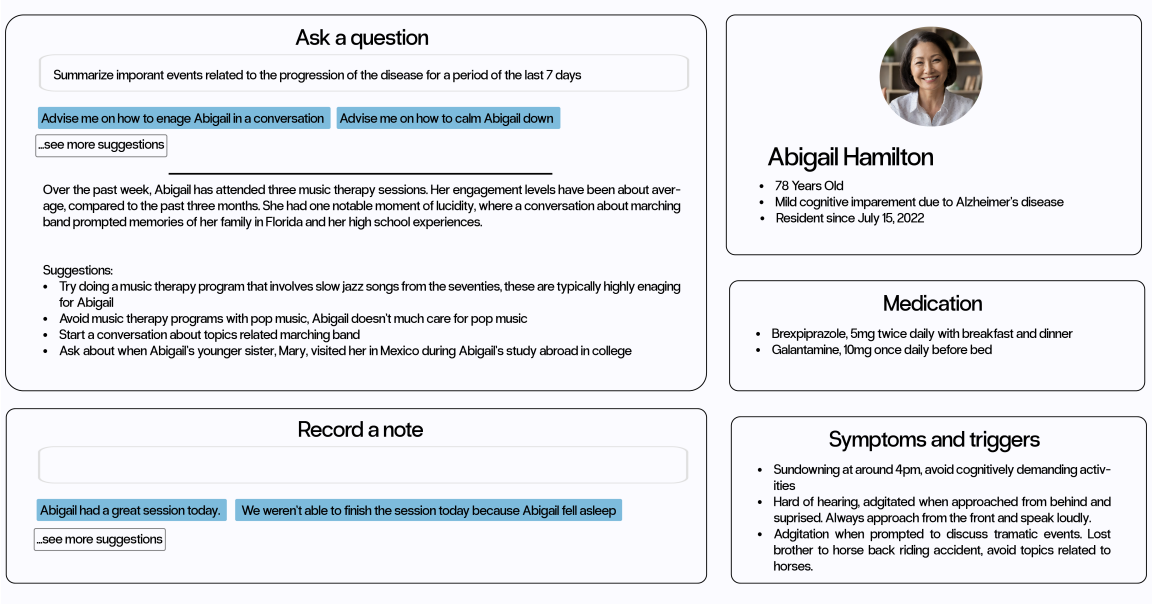


Figure C.2: Excerpt from medium-fidelity prototypes. All user content shown here is mocked data.

C.2 Detailed System Overview

C.2.1 Onboarding

FITL includes an onboarding page featuring walkthrough videos that I created to guide users through each view. First-time users are automatically directed to this onboarding page upon opening the app. No account creation or manual data entry is required; user identity, role, and care recipient information are seamlessly integrated through single sign-on via Memcara's existing platform.

C.2.2 Reels View

Reels View serves as the primary entry point for families, reflecting their preference for regular, emotionally meaningful updates. Updates come in the form of short video clips, or "reels", drawn from structured therapeutic interactions between staff and residents, and

they provide a low-effort way to remain connected to the care experience, particularly when in-person observation is not feasible.

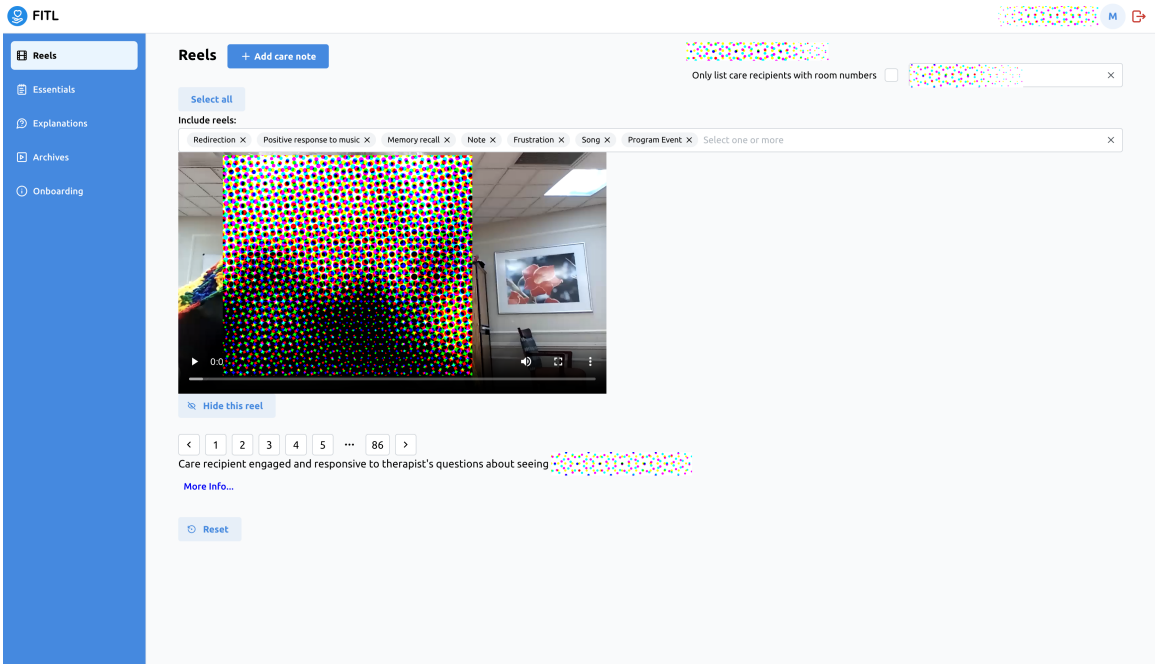


Figure C.3: A screenshot of Reels View. All personally identifiable information has been pixelated.

Video capture for these 1:1 sessions occurs passively by leveraging the Memcara platform's existing tooling. Caregivers use a fixed-position iPad[1] already in place for delivering assistive recorded content (e.g., music videos with generic memory related question prompts) during therapy—which continuously faces the care recipient, allowing sessions to be captured with minimal effort. Caregivers can initiate video capture with a single tap, and the device reliably records a well-framed interaction without requiring additional setup or repositioning. This design minimizes the burden on care staff, who can document meaningful moments in ways that align with their existing workflows and without disrupting the flow of the session.

While many reels highlight positive moments—such as a resident singing along to music, recognizing a photograph, or laughing during a conversation—Reels View may also

[1]Currently, the partner facilities have only a few iPads that are shared by the staff, these are repositioned each time a 1-1 staff–care recipient interaction takes place.

surface more difficult or complex episodes, such as a moment of confusion redirected effectively by a caregiver or a resident expressing frustration during a memory activity. These moments are included when they are marked by atypical behaviors or salient caregiver responses, offering families greater transparency into the realities of dementia care while supporting reflection and learning. Reels are dynamically selected by the CareInsights backend.

Each reel is paired with an AI-generated summary synthesized by the CareInsights backend, such as "Resident identifies their daughter in a photo and talks about a family trip," "Initially resists participating in memory exercise but begins to engage after caregiver prompts with familiar music," or "Caregiver calmly redirects agitation during afternoon confusion episode." These captions aim to foreground the emotional or narrative significance of the moment without requiring families to interpret raw footage. The "More Info" button beneath each summary transitions the user to Explanations View (see Section C.2.3), where the system pre-populates a clarifying question (e.g., "What prompted the change in mood during the therapy session on <Date:Time>?") and returns an explanation grounded in the surrounding care record.

To help families manage emotional load, Reels can be filtered using a horizontal bar that groups clips by topic or emotional tone—for instance, "Positive response to music," "Redirection," or "Memory recall." This allows families to avoid—or stay updated on—content that may be emotionally difficult. Individual clips can also be hidden, and doing so updates the backend's clip selection algorithm, deprioritizing similar moments in future Reels by updating the backend curation logic.

Reels View is embedded within a shared data infrastructure designed to accommodate iterative feedback. During prototyping, it was found that families' responses to clips varied widely—not only in emotional interpretation, but in what was considered helpful or appropriate to share. Reels includes an "Add Care Note" button, allowing users to comment on moments with reflections or clarifications (e.g., *"He always did love Kathleen. He used*

to hum her favorite song, Vienna, when he missed her."). These notes are appended to the CareInsights data layer and trigger re-synthesis of the AI-generated summary, influencing both the displayed caption, the system's evolving model of which moments to surface in the future, and AI-generated data on other views.

Reels is intentionally lightweight. The interface presents a paginated feed of video clips with thumbnails emphasizing facial expressions or gestures. All content is drawn from structured therapy contexts—not continuous surveillance—and administrative controls in Archives View allow staff to review or restrict what recordings can be shared at a per-session level[2].

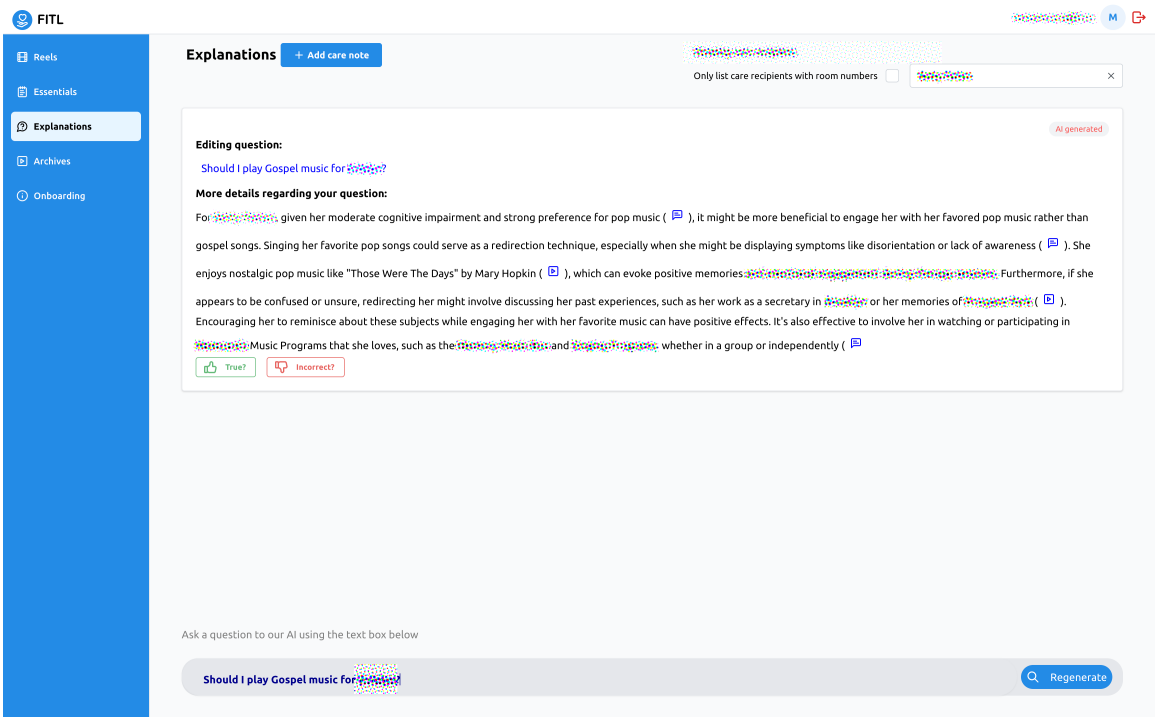


Figure C.4: Explanations View. All personally identifiable information has been pixelated.

C.2.3 Explanations View

The Explanations View enables both families and staff to request clarification, context, or elaboration on care-related observations. These questions may pertain to recent interactions

[2]Placing all privacy controls in the Archives View was, conveniently, both more manageable for staff in practice and allows for more flexibility when making dynamic content updates.

(e.g., “Describe the most recent care activities”), behavioral patterns (“Has he been showing signs of agitation now more than usual?”), or care practices (“What usually helps bring out moments of lucidity?”).

When a user enters the view—either via a drill-down link or directly from the navigation sidebar—they are presented with a suggested question and answer based on the preceding interaction, meaning the view is never without content[3]. The existing prompt can then be updated.

To promote transparency without overwhelming users, FITL’s response to the query appears as a short paragraph with the user’s question at the top. Reference icons embedded in the text link directly to supporting documentation, such as text-based notes (i.e., from EHR systems or staff/family feedback) (💬) or timestamped video clips (▶). For example: *“Caregiver observed increased agitation, music therapy involving Andrea Bocelli was used as a calming technique.”* (▶) Clicking on these citations helps users do a deep dive into the underlying data, to verify how a conclusion was reached.

Like the real-world practices it emulates—such as phone calls or in person exchanges—Explanations View supports a conversational interaction model. Users can submit follow-up or rephrased questions, prompting the system to regenerate responses using updated query parameters. This enables asynchronous, lightweight dialogue that mirrors the kinds of clarifications families might seek during visits, or that staff routinely exchange as part of everyday coordination.

Like Reels and Essentials, Explanations View includes an “Add Care Note” button for free response. However, at users’ request for a faster way to validate or reject responses for the system to learn, corresponding “True?” and “Incorrect?” buttons were also included. On the backend, feedback is tied to the underlying records and prompts the system to reprocess its explanation (and other views that reference the updated info) accordingly.

[3] In early prototyping, it was found that users got confused by or unsure how to use views that felt “empty”; having a default question and answer showing at all times gave users some context for both how to use the view as well as mentally triggered them to think of additional questions to ask.

For example, a family member might write, “That’s not a good redirection strategy for Lucy—she responds better to music and mentions of her sister, Rachel.” This input is saved to the CareInsights data layer and influences future outputs involving the resident. A similar (though, less precise[4] result) could be achieved by clicking on the “Incorrect?” button. Over time, these mechanisms help refine the system’s understanding of individual preferences and embed practical care strategies into the shared knowledge base.

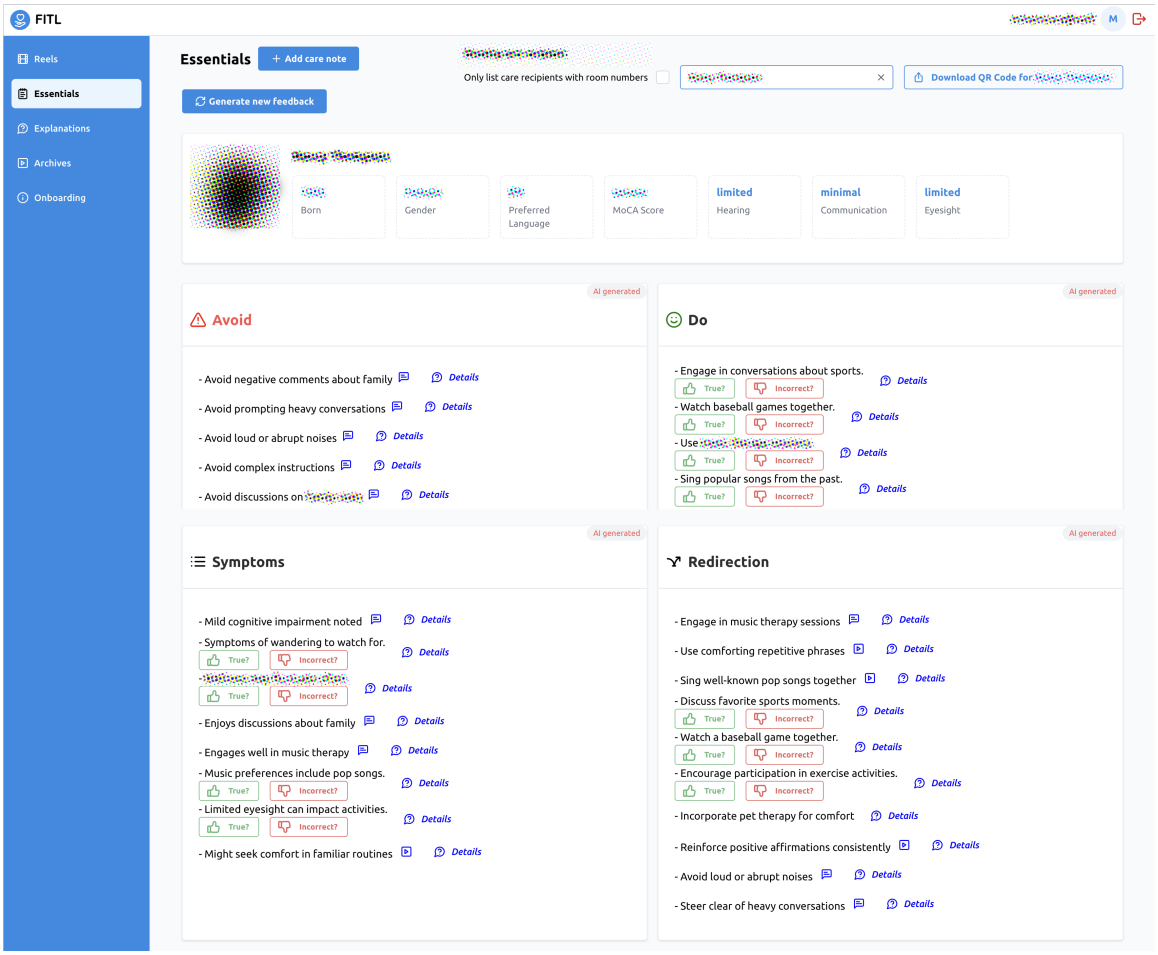


Figure C.5: A screenshot of Essentials View. All personally identifiable information has been pixelated.

[4]While the initial response would be updated to avoid making the noted claim, without the added context of a free response input, the new result could take a more generic sound. For this reason, users took the time to use the “Add Care Note” button, encouraged by the sense that FITL was learning.

C.2.4 Essentials View

The Essentials View was created after seeing how staff often carried informal "cheat sheets" summarizing a resident's most important traits—such as how they tend to behave on a typical day, what strategies can defuse their agitation, and which activities spark engagement or cause distress. By transforming those hand-scrawled bullet points into a digitally maintained list, this helps ensure that every member of the care team, as well as family members with the right permissions, can see the same up-to-date guidance. Following staff members' suggestions, Essentials View is split into five boxes:

1. A top-level, "Profile" box with the most important info (i.e., name, preferred language, communication capabilities, hearing capabilities, etc.),
2. "⚠ Avoid" (i.e., activities or topics to avoid),
3. "☺ Do" (i.e., activities or topics to engage in),
4. "⁝≡ Symptoms" (i.e., common dementia-related behaviors and comorbidities exhibited by this individual), and
5. "⅄ Redirection" (i.e., effective strategies for redirecting the individual).

The four bottom boxes house a brief set of bullet points summarizing insights that are identified as relevant to delivering person-centered care.

From a technical perspective, this is handled by querying the backend with a corresponding question—"What behaviors are typical?", "How do we de-escalate them?", "What helps them engage?", "What should we avoid?"—and specifying that the format of the response should be a bulleted list. The backend pipeline gathers data relevant (i.e., processed text summaries of videos, EHR system data, and user feedback) to each question, synthesizes a response, and then formats that response into the requested bulleted format. The template passed into the API specifically limits the word count per bullet, so that staff, or family, can understand the most important takeaways at a glance. This is the same API

used for generating responses on Explanations View, however, in that case a different template is used to achieve a longer, narrative form response. As with Explanations View, claims here are supported by citations that function as transitions to Archives View.

A "Details" button accompanying each bullet was included at the request of staff who wanted a way to quickly navigate to the Explanations View and pre-populate a related follow-up question to the bullet (i.e., "tell me more about <bullet content>"). This was important because bullets are necessarily limited in their ability to convey nuance, and most users expressed a need for an (ideally, "one-click") option to get more "details" on demand.

To address moments of system uncertainty, a minimal feedback mechanism was included: thumbs-up and thumbs-down icons appear next to certain bullet points in the Essentials View. These icons are not shown universally. Rather, they appear only when the system detects that a bullet lacks explicit citations—usually in cases where few structured records exist to back the claim[5]. This approach emerged from user feedback: staff and families wanted a way to quickly respond to suggestions that felt too generic or possibly misleading, but they also recognized that some of these uncited suggestions were helpful and should not be removed outright.

The presence of thumbs icons signals that the system itself considers the statement provisional. Users can click thumbs-up or thumbs-down and optionally add a brief comment. If approved, a system-generated note is stored in the data layer (e.g., "Caregiver [Name] confirmed this strategy on [Date]"), which serves as a supporting record. The item is then considered "certain" and the icons are removed in future views. This balances the risk of misleading content with the potential value of AI-generated suggestions, especially for care recipients with limited data on record.

[5]In practice, it was found that bullets without citations only appear for care recipients with very few recorded videos (e.g., fewer than ten, though this can vary). In these cases, the generated content—but particularly uncertain bullet points—tends to be generic and may be less helpful or even misleading. Once a care recipient has more than ten videos on record, the content in this view becomes more grounded in observed events and behavioral patterns. At that point, offering quick feedback becomes less appropriate, as edits typically require more nuanced reasoning and justification.

In contrast, when a bullet is already tied to an existing care note or transcript reference, errors require more than a vote of confidence to resolve. These cases demand a higher degree of accountability: the underlying record must be corrected or augmented. To support this, users are directed to the "Add Care Note" button for submitting corrections. Requiring written input in these cases ensures that edits are well-documented, supporting downstream data quality across all views.

When feedback is submitted, the system updates immediately, allowing users to see how their input was interpreted. If the update fails to address the issue, the visible output provides a concrete prompt for refinement—another care note can be added to clarify or correct what the system inferred. This iterative model was observed to be critical to maintaining data accuracy, while also making system behavior transparent. In more involved cases, users can navigate to the source record and directly edit it—ensuring that the system remains grounded in a reviewable, collectively maintained data layer.

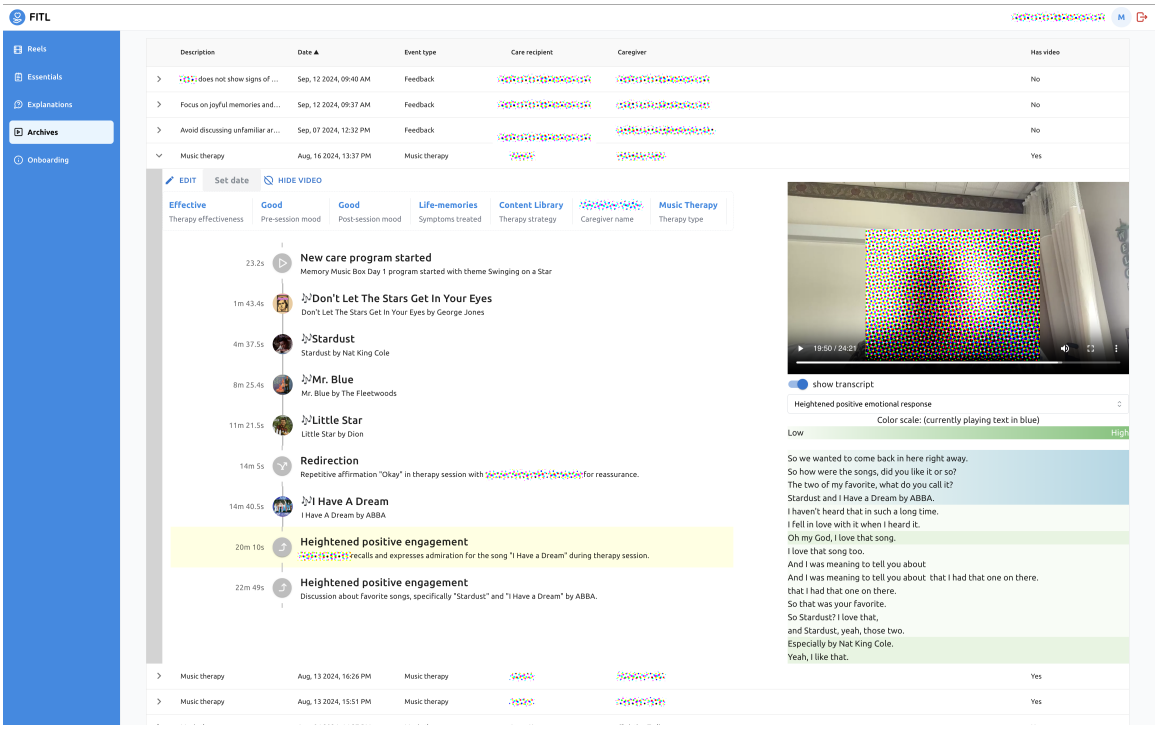


Figure C.6: A screenshot of Archives View. The header and the "Add Care Note" button are not shown (to capture more of the other features in-frame), but appear at the top of the page. All personally identifiable information has been pixelated.

C.2.5 Archives View

Archives View serves as the canonical, full-fidelity interface for interacting with a resident's longitudinal care record. Inspired by the paper-based tracking systems observed among "super-user" facility staff, the view was designed to replicate the format of hand-curated binders that included timelines, annotations, and program notes. This design—vertical, expandable, and chronologically ordered—offers an at-a-glance overview of recent care and the option to drill down into fine-grained documentation. It functions in FITL as a centralized reference interface, supporting both record verification and the editing of underlying care records that drive updates across the rest of the views (which show dynamically generated content).

To support these needs, Archives View displays both structured and multimedia documentation. Text-only records (e.g., shift logs, staff notes, family-submitted observations) appear in the expandable list alongside video records. When a record is derived from a video recording, a set of embedded subviews becomes available (see Figure C.6): the video is shown alongside a transcript, with dynamic highlighting based on AI-generated annotations. These annotations identify caregiver-relevant behaviors like memory recall, emotional engagement, or behavioral deviations. Highlighting can be toggled to emphasize different engagement types, and clicking on transcript lines or timeline markers jumps the video to the associated moment.

Adjacent to the transcript, a vertically stacked timeline visualizes system-detected moments of salience—such as caregiver responses, family-relevant clips (used in Reels View), and successful redirection. When music is detected, album art appears inline. Like the transcript, the timeline serves as an interactive index for navigating the video. It highlights the current timestamp and allows users to jump to specific moments with a single click. On the backend, the timeline is generated using timestamped logs from Memcara's platform, which provides a continuous stream of events corresponding to stimuli presented during therapy sessions—such as images, music, or verbal prompts. Because the platform

records video concurrently with this structured event stream, the system can align what the care recipient is seeing or hearing at any given moment with how they behave or respond in the footage. This integration enables the system to reason about potential causal relationships—for example, identifying whether an observed reaction (such as memory recall or emotional shift) coincided with a specific song or visual prompt. These aligned timelines allow caregivers and families to more easily understand not just what happened, but why—supporting clinical interpretation, family insight, and more targeted care planning. Taken together, these elements offer a comprehensive picture of each recorded session, dramatically reducing the time required to understand what happened—and solving a common complaint raised by staff during early prototyping: no one had time to watch entire videos.

Editing tools in Archives View are designed to support both granular and systemic updates. Staff can revise the AI-generated transcript (e.g., correcting a misinterpreted word), edit or re-label a record's metadata, or add entirely new entries. These updates propagate across Reels, Essentials, and Explanations Views, ensuring that all representations of a resident's care remain consistent. Because all generated content in FITL maintains references to its source records—either explicitly or internally—the system uses targeted re-rendering to update only content affected by a change.

This propagation logic—driven by dependency-tracking and view-aware update rules—ensures that even lightweight feedback (like a thumbs-down click) can cascade across views[6]. In cases where ambiguity arises (i.e., "Add a Care Note" on Essentials View), page context (i.e., what citations are present) helps narrow down which record is most likely being referenced, enabling even brief notes to be actionable. Manual edits remain visible as part of the record's revision history, making feedback legible and accountable.

[6]Feedback submitted through actions like "Thumbs-down" or "Add a Care Note" is recorded as a new entry in Archives View and linked to the specific record it references—forming a traceable graph of relationships. This structure enables FITL to backpropagate updates that originate as feedback on other views (rather than direct edits on Archives View), allowing it to revise dependent content while maintaining a transparent and traceable history of changes.

Visibility controls were designed to be intentionally simple. Staff can set default permissions—e.g., defaulting new content to "Staff-only" or "Family-visible"—or adjust them record by record. During late-stage prototyping, two partner sites informed the team that EHR-linked media and transcripts were sometimes subject to "on-prem only" access policies. In response, geofencing was added: when enabled, the system checks the user's GPS location before loading restricted records. This feature provided staff with additional reassurance that sensitive content would remain within facility bounds when required, balancing accountability and protection.

Archives View serves as the foundation of the FITL system, providing a persistent and editable record of care from which all other views derive their content. While other views offer synthesized and filtered perspectives, Archives gives users full access to the raw data and tools to shape what the system learns. By combining editable records, media navigation, and targeted propagation of updates, Archives promotes a high-trust documentation experience that reflects the constraints, values, and practices of long-term dementia care.

Bibliography

David J. Abson, Joern Fischer, Julia Leventon, Jens Newig, Thomas Schomerus, Ulli Vilsmaier, Henrik von Wehrden, Paivi Abernethy, Christopher D. Ives, Nicolas W. Jager, and Daniel J. Lang. Leverage points for sustainability transformation. *Ambio*, 46(1):30–39, 2017. doi: 10.1007/s13280-016-0800-y.

Mark S. Ackerman and Christine A. Halverson. Organizational memory as objects, processes, and trajectories: An examination of organizational memory in use. *Computer Supported Cooperative Work (CSCW)*, 13(2):155–189, 2004. doi: 10.1023/B:COSU.0000045805.77534.2a.

Mark S. Ackerman, Juri Dachtera, Volkmar Pipek, and Volker Wulf. Sharing knowledge and expertise: The CSCW view of knowledge management. *Computer Supported Cooperative Work (CSCW)*, 22(4–6):531–573, 2013. doi: 10.1007/s10606-013-9192-8.

Nancy E Adams. Bloom's taxonomy of cognitive learning objectives. *Journal of the Medical Library Association: JMLA*, 103(3):152–153, 2015. doi: 10.3163/1536-5050.103.3.010.

Molla Imaduddin Ahmed, Brendan Spooner, John Isherwood, Mark Lane, Emma Orrock, and Ashley Dennison. A systematic review of the barriers to the implementation of artificial intelligence in healthcare. *Cureus*, 15(10):e46454, 2023. doi: 10.7759/cureus.46454.

Yusuf Ahmed, Martin Ferguson-Pell, Kim Adams, and Adriana Ríos Rincón. EEG-based engagement monitoring in cognitive games. *Sensors*, 25(7):2072, 2025. doi: 10.3390/s25072072.

AI4ALL. Open learning curriculum. `https://ai-4-all.org/resources/`, 2023. Page no longer live; archived at `https://web.archive.org/web/20231124084446/https://ai-4-all.org/resources/` (snapshot 2023-11-24). Accessed 2026-08-03.

AI4K12 Initiative. AI4K12 grade band progression charts. `https://ai4k12.org/gradeband-progression-charts/`, 2020. Working drafts unpacking the Five Big Ideas in AI into grade-band learning objectives and "Unpacked" descriptions. Accessed 2026-08-19.

Mary Ainley, Suzanne Hidi, and Dagmar Berndorff. Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology*, 94(3): 545–561, 2002. doi: 10.1037/0022-0663.94.3.545.

Isabelle Rose I Alberto, Nicole Rose I Alberto, Arnab K Ghosh, Bhav Jain, Shruti Jayakumar, Nicole Martinez-Martin, Ned McCague, Dana Moukheiber, Lama Moukheiber, Mira Moukheiber, Sulaiman Moukheiber, Antonio Yaghy, Andrew Zhang, and Leo Anthony Celi. The impact of commercial health datasets on medical research and healthcare algorithms. *The Lancet Digital Health*, 5(5):e288–e294, 2023. doi: 10.1016/S2589-7500(23)00025-0.

John Alford and Brian W. Head. Wicked and less wicked problems: A typology and a contingency framework. *Policy and Society*, 36(3):397–413, 2017. doi: 10.1080/14494035.2017.1361634.

Abdullah Alismail, Brett Schaeffer, Andrea Oh, Saba Hamiduzzaman, Noha Daher, Hae-Young Song, Brian Furukawa, and Laren D Tan. The use of the net promoter score (NPS) in an outpatient allergy and pulmonary clinic: an innovative look into using tablet-based tool vs traditional survey method. *Patient Related Outcome Measures*, 11:137–142, 2020. doi: 10.2147/PROM.S248431.

Mohammed Abdullatif Almulla. The effectiveness of the project-based learning (PBL) approach as a way to engage students in learning. *Sage Open*, 10(3):2158244020938702, 2020. doi: 10.1177/2158244020938702.

Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–13, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450359702. doi: 10.1145/3290605.3300233. URL `https://doi.org/10.1145/3290605.3300233`.

Janet Carter Anand, Gavin Davidson, Berni Kelly, and Geraldine Macdonald. Personalization of care: A wicked problem or a wicked solution? In Will Thomas, Anneli Hujala, Sanna Laulainen, and Robert McMurray, editors, *The Management of Wicked Problems in Health and Social Care*, pages 34–48. Routledge, 2018. doi: 10.4324/9781315102597-4.

Kanthee Anantapong, Nathan Davies, and Elizabeth L Sampson. Communication between the multidisciplinary team and families regarding nutrition and hydration for people with severe dementia in acute hospitals: a qualitative study. *Age and Ageing*, 51(11):afac230, 2022. doi: 10.1093/ageing/afac230.

Amirhossein Eslami Andargoli, Nalika Ulapane, Tuan Anh Nguyen, Nadeem Shuakat, John Zelcer, and Nilmini Wickramasinghe. Intelligent decision support systems for dementia care: A scoping review. *Artificial Intelligence in Medicine*, 150:102815, 2024. doi: 10.1016/j.artmed.2024.102815.

Lorin W. Anderson and David R. Krathwohl, editors. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman, New York, complete edition, 2001. ISBN 0321084055.

George L Anesi and Meeta Prasad Kerlin. The impact of resource limitations on care delivery and outcomes: routine variation, the coronavirus disease 2019 pandemic, and persistent shortage. *Current Opinion in Critical Care*, 27(5):513–519, 2021. doi: 10.1097/MCC.0000000000000859.

James M Applefield, Richard Huber, and Mahnaz Moallem. Constructivism in theory and practice: Toward a better understanding. *The High School Journal*, 84(2):35–53, 2000.

Aristotle. *Poetics*. Penguin Books, London, 1996. Translated by Malcolm Heath. Originally composed c. 335 BCE.

Ross D. Arnold and Jon P. Wade. A definition of systems thinking: A systems approach. *Procedia Computer Science*, 44:669–678, 2015. doi: 10.1016/j.procs.2015.03.050.

Sarah L Ash and Patti H Clayton. Generating, deepening, and documenting learning: The power of critical reflection in applied learning. *Journal of Applied Learning in Higher Education*, 1:25–48, 2009. doi: 10.57186/jalhe_2009_v1a2p25-48.

Ramona Backhaus, Linda JM Hoek, Erica de Vries, Jolanda CM van Haastregt, Jan PH Hamers, and Hilde Verbeek. Interventions to foster family inclusion in nursing homes for people with dementia: a systematic review. *BMC Geriatrics*, 20(1):434, 2020. doi: 10.1186/s12877-020-01836-w.

Abeer Badawi, Somayya Elmoghazy, Samira Choudhury, Sara Elgazzar, Khalid Elgazzar, and Amer M Burhan. Multimodal detection of agitation in people with dementia in clinical settings: Observational pilot study. *JMIR Aging*, 8:e68156, 2025. doi: 10.2196/68156.

Amee Baird, Olivia Brancatisano, Rebecca Gelding, and William Forde Thompson. Characterization of music and photograph evoked autobiographical memories in people with Alzheimer's disease. *Journal of Alzheimer's Disease*, 66(2):693–706, 2018. doi: 10.3233/JAD-180627.

Clive G Ballard, Serge Gauthier, Jeffrey L Cummings, Henry Brodaty, George T Grossberg, Philippe Robert, and Constantine G Lyketsos. Management of agitation and aggression associated with Alzheimer disease. *Nature Reviews Neurology*, 5(5):245–255, 2009. doi: 10.1038/nrneurol.2009.39.

Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel S. Weld. Does the whole exceed its parts? The effect of AI explanations on complementary team performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, pages 1–16, New York, NY, USA, 2021. Association for Computing Machinery. doi: 10.1145/3411764.3445717.

Jørgen P. Bansler, Erling C. Havn, Kjeld Schmidt, Troels Mønsted, Helen Høgh Petersen, and Jesper Hastrup Svendsen. Cooperative epistemic work in medical practice: an analysis of physicians' clinical notes. *Computer Supported Cooperative Work (CSCW)*, 25 (6):503–546, 2016. doi: 10.1007/s10606-016-9261-x.

Qiming Bao, Juho Leinonen, Alex Yuxuan Peng, Wanjun Zhong, Gaël Gendron, Timothy Pistotti, Alice Huang, Paul Denny, Michael Witbrock, and Jiamou Liu. Exploring iterative enhancement for improving learnersourced multiple-choice question explanations with large language models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(28):28955–28963, 2025. doi: 10.1609/aaai.v39i28.35164. Published version of arXiv:2309.10444, previously circulated as "Exploring Self-Reinforcement for Improving Learnersourced Multiple-Choice Question Explanations with Large Language Models".

Rachel Barken and Ruth Lowndes. Supporting family involvement in long-term residential care: Promising practices for relational care. *Qualitative Health Research*, 28(1):60–72, 2018. doi: 10.1177/1049732317730568.

Caroline A. Bartel and Raghu Garud. The role of narratives in sustaining organizational innovation. *Organization Science*, 20(1):107–117, 2009. doi: 10.1287/orsc.1080.0372.

Amna Basharat. Learnersourcing thematic and inter-contextual annotations from islamic texts. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 92–97, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450340823. doi: 10.1145/2851581.2890386.

Michael Bauer and Rhonda Nay. Improving family–staff relationships in assisted living facilities: the views of family. *Journal of Advanced Nursing*, 67(6):1232–1241, 2011. ISSN 0309-2402. doi: 10.1111/j.1365-2648.2010.05575.x.

Zeinab Bedri, Ruairí de Fréin, and Geraldine Dowling. Community-based learning: A primer. *Irish Journal of Academic Practice*, 6(1):5, 2017. doi: 10.21427/D7F143.

Brian R. Belland, Andrew E. Walker, Nam Ju Kim, and Mason Lefler. Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. *Review of Educational Research*, 87(2):309–344, 2017. doi: 10.3102/0034654316670999.

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, pages 610–623. ACM, 2021. doi: 10.1145/3442188.3445922.

Joshua M. Bentley. Sensemaking theory. In Craig E. Carroll, editor, *The SAGE Encyclopedia of Corporate Reputation*, volume 2, pages 743–744. SAGE Publications, Thousand Oaks, CA, 2016. doi: 10.4135/9781483376493.n277.

Marc Berg and Geoffrey Bowker. The multiple bodies of the medical record: Toward a sociology of an artifact. *The Sociological Quarterly*, 38(3):513–537, 1997. doi: 10.1111/j.1533-8525.1997.tb00490.x.

Matthew L. Bernacki, Timothy J. Nokes-Malach, and Vincent Aleven. Examining self-efficacy during learning: Variability and relations to behavior, performance, and learning. *Metacognition and Learning*, 10(1):99–117, 2015. doi: 10.1007/s11409-014-9127-x.

Karabi C. Bezboruah, Darla Paulson, and Jason Smith. Management attitudes and technology adoption in long-term care facilities. *Journal of Health Organization and Management*, 28(3):344–365, 2014. doi: 10.1108/JHOM-11-2011-0118.

Karthik S. Bhat, Mohit Jain, and Neha Kumar. Infrastructuring telehealth in (In)Formal patient-doctor contexts. *Proceedings of the ACM on Human-Computer Interaction*, 5 (CSCW2):1–28, 2021. doi: 10.1145/3476064.

Karthik S. Bhat, Amanda K. Hall, Tiffany Kuo, and Neha Kumar. "We are half-doctors": Family caregivers as boundary actors in chronic disease management. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1):1–29, 2023. doi: 10.1145/3579545.

Kiran Bisra, Qing Liu, John C. Nesbit, Farimah Salimi, and Philip H. Winne. Inducing self-explanation: A meta-analysis. *Educational Psychology Review*, 30(3):703–725, 2018. doi: 10.1007/s10648-018-9434-x.

Courtney K. Blackwell, Alexis R. Lauricella, and Ellen Wartella. The influence of TPACK contextual factors on early childhood educators' tablet computer use. *Computers & Education*, 98:57–69, 2016. doi: 10.1016/j.compedu.2016.02.010.

Benjamin S. Bloom, Max D. Engelhart, Edward J. Furst, Walker H. Hill, and David R. Krathwohl. *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*. David McKay Company, New York, 1956.

Phyllis C. Blumenfeld, Elliot Soloway, Ronald W. Marx, Joseph S. Krajcik, Mark Guzdial, and Annemarie Palincsar. Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist*, 26(3-4):369–398, 1991. doi: 10.1080/00461520.1991.9653139.

Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray

Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. On the opportunities and risks of foundation models, 2021. arXiv preprint arXiv:2108.07258.

Charles C. Bonwell and James A. Eison. Active learning: Creating excitement in the classroom. ASHE-ERIC Higher Education Report 1, The George Washington University, School of Education and Human Development, Washington, DC, 1991. URL `https://eric.ed.gov/?id=ED336049`. ERIC Document ED336049.

Claus Bossen and Randi Markussen. Infrastructuring and ordering devices in health care: Medication plans and practices on a hospital ward. *Computer Supported Cooperative Work (CSCW)*, 19(6):615–637, 2010. doi: 10.1007/s10606-010-9131-x.

Alexios Brailas. Rhizomatic learning in action: A virtual exposition for demonstrating learning rhizomes. In *Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality*, TEEM'20, pages 309–314, New York, NY, USA, 2020. Association for Computing Machinery. doi: 10.1145/3434780.3436565.

Teresa Brannick and David Coghlan. In defense of being "Native": The case for insider academic research. *Organizational Research Methods*, 10(1):59–74, 2007. doi: 10.1177/1094428106289253.

Lindsey Brett, Victoria Traynor, and Paul Stapley. Effects of physical exercise on health and well-being of individuals living with a dementia in nursing homes: a systematic review. *Journal of the American Medical Directors Association*, 17(2):104–116, 2016. doi: 10.1016/j.jamda.2015.08.016.

Nathan R. Bromberg, Angsana A. Techatassanasoontorn, and Antonio Díaz Andrade. Engaging students: Digital storytelling in information systems learning. *Pacific Asia Journal of the Association for Information Systems*, 5(1):1–22, 2013. doi: 10.17705/1pais.05101.

Dawn Brooker. What is person-centred care in dementia? *Reviews in Clinical Gerontology*, 13(3):215–222, 2003. doi: 10.1017/S095925980400108X.

Dawn Brooker and Isabelle Latham. *Person-Centred Dementia Care: Making Services Better with the VIPS Framework*. Jessica Kingsley Publishers, London, second edition, 2015. ISBN 9781849056663.

Ruth Brookman, Simon Parker, Leonard Hoon, Asuka Ono, Atsushi Fukayama, Hisashi Matsukawa, and Celia B. Harris. Technology for dementia care: what would good technology look like and do, from carers' perspectives? *BMC Geriatrics*, 23(1):867, 2023. doi: 10.1186/s12877-023-04530-9.

John Seely Brown and Paul Duguid. Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation. *Organization Science*, 2(1): 40–57, 1991. doi: 10.1287/orsc.2.1.40.

Sonya Brownie and Susan Nancarrow. Effects of person-centered care on residents and staff in aged-care facilities: a systematic review. *Clinical Interventions in Aging*, 8: 1–10, 2013. doi: 10.2147/CIA.S38589.

Jerome Bruner. The narrative construction of reality. *Critical Inquiry*, 18(1):1–21, 1991. doi: 10.1086/448619.

Jerome Seymour Bruner. *Actual Minds, Possible Worlds*. Harvard University Press, Cambridge, MA and London, England, 1986. ISBN 9780674029019. doi: 10.4159/9780674029019. URL https://doi.org/10.4159/9780674029019.

Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. Sparks of Artificial General Intelligence: Early experiments with GPT-4, 2023. URL https://arxiv.org/abs/2303.12712. arXiv:2303.12712.

Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–21, 2021. doi: 10.1145/3449287.

Nanci M. Burk. Empowering at-risk students: Storytelling as a pedagogical tool, 2000. URL https://eric.ed.gov/?id=ED447497. Paper presented at the Annual Meeting of the National Communication Association (86th, Seattle, WA, November 9–12, 2000). ERIC Document ED447497. Accessed: 2026-08-03.

Marie-Andrée Cadieux, Linda J Garcia, and Jonathan Patrick. Needs of people with dementia in long-term care: A systematic review. *American Journal of Alzheimer's Disease & Other Dementias®*, 28(8):723–733, 2013. doi: 10.1177/1533317513500840.

Andrea Capocci, Vito D. P. Servedio, Francesca Colaiori, Luciana S Buriol, Debora Donato, Stefano Leonardi, and Guido Caldarelli. Preferential attachment in the growth of social networks: The internet encyclopedia Wikipedia. *Physical Review E*, 74(3): 036116, 2006. doi: 10.1103/PhysRevE.74.036116.

Ashley Capoot. OpenAI cyber models broke out of training environment to hack Hugging Face. CNBC, 2026. URL https://www.cnbc.com/2026/07/22/open-ai-cyber-models-hack-hugging-face.html. Published 2026-07-22; reporting on OpenAI's disclosure that its models autonomously escaped a sandboxed evaluation and exploited a vulnerability to breach Hugging Face. Accessed: 2026-08-03.

Áine Carroll, Claire Collins, Jane McKenzie, Diarmuid Stokes, and Andrew Darley. Application of complexity theory in health and social care research: a scoping review. *BMJ Open*, 13(3):e069180, 2023. doi: 10.1136/bmjopen-2022-069180.

Mary W Chaffee and Margaret M McNeill. A model of nursing as a complex adaptive system. *Nursing Outlook*, 55(5):232–241, 2007. doi: 10.1016/j.outlook.2007.04.003.

Rita Charon. Narrative medicine: A model for empathy, reflection, profession, and trust. *JAMA*, 286(15):1897–1902, 2001. doi: 10.1001/jama.286.15.1897.

You Chen, Christoph U Lehmann, and Bradley Malin. Digital information ecosystems in modern care coordination and patient care pathways and the challenges and opportunities for AI solutions. *Journal of Medical Internet Research*, 26:e60258, 2024. doi: 10.2196/60258.

Josh Cherian, Samantha Ray, Thomas Mernar, Paul Taele, Helen Mach, Jung In Koh, Ping Ye, and Tracy Hammond. A step toward better care: Understanding what caregivers and residents in assisted living facilities value in health monitoring systems. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1):1–29, 2024. doi: 10.1145/3637290.

Harun Cigdem and Mustafa Ozturk. Critical components of online learning readiness and their relationships with learner achievement. *Turkish Online Journal of Distance Education*, 17(2):98–109, 2016. doi: 10.17718/tojde.09105.

Camilla N Clark and Jason D Warren. Music, memory and mechanisms in Alzheimer's disease. *Brain*, 138(8):2122–2125, 2015. doi: 10.1093/brain/awv148.

Donald Clark. Bloom (1913–1999) – Mastery learning. Taxonomy of learning... not a hierarchy.... Donald Clark Plan B (blog), `https://donaldclarkplanb.blogspot.com/2020/03/bloom-1913-1999-taxonomy-of-learning.html`, March 2020. Accessed 2026-08-03.

Allan Collins, John Seely Brown, and Susan E. Newman. Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In Lauren B. Resnick, editor, *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*, pages 453–494. Lawrence Erlbaum Associates, Hillsdale, NJ, 1989.

Tina Comes. Sensemaking AI: Introducing a research and design agenda for human–AI networks. *EPJ Data Science*, 15(1):42, 2026. doi: 10.1140/epjds/s13688-026-00634-5.

Tina Comes, Bartel Van de Walle, and Luk Van Wassenhove. The coordination-information bubble in humanitarian response: Theoretical foundations and empirical investigations. *Production and Operations Management*, 29(11):2484–2507, 2020. doi: 10.1111/poms.13236.

AnneMarie M Conley. Patterns of motivation beliefs: Combining achievement goal and expectancy-value perspectives. *Journal of Educational Psychology*, 104(1):32–47, 2012. doi: 10.1037/a0026042.

Gregorio Convertino, Anna Wu, Xiaolong Zhang, Craig H. Ganoe, Blaine Hoffman, and John M. Carroll. Designing group annotations and process visualizations for role-based collaboration. In *Social Computing, Behavioral Modeling, and Prediction*, pages 197–206. Springer, Boston, MA, 2008. doi: 10.1007/978-0-387-77672-9_22.

Robert J. Crutcher and Alice F. Healy. Cognitive operations and the generation effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(4):669–675, 1989. doi: 10.1037/0278-7393.15.4.669.

Lola L. Cuddy, Ritu Sikka, Kristen Silveira, Sean Bai, and Ashley Vanstone. Music-evoked autobiographical memories (MEAMs) in Alzheimer disease: Evidence for a positivity effect. *Cogent Psychology*, 4(1):1277578, 2017. doi: 10.1080/23311908.2016.1277578.

Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. Wizard of Oz studies—why and how. In *Proceedings of the 1st International Conference on Intelligent User Interfaces*, IUI '93, pages 193–200, New York, NY, USA, 1993. Association for Computing Machinery. doi: 10.1145/169891.169968.

Ali Darvishi, Hassan Khosravi, and Shazia Sadiq. Employing peer review to evaluate the quality of student generated content at scale: A trust propagation approach. In *Proceedings of the Eighth ACM Conference on Learning @ Scale*, L@S '21, pages 139–150, New York, NY, United States, 2021. Association for Computing Machinery. doi: 10.1145/3430895.3460129.

Vedant Das Swain, Qiuyue "Joy" Zhong, Jash Rajesh Parekh, Yechan Jeon, Roy Zimmermann, Mary P. Czerwinski, Jina Suh, Varun Mishra, Koustuv Saha, and Javier Hernandez. AI on my shoulder: Supporting emotional labor in front-office roles with an LLM-based empathetic coworker. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–29, New York, NY, USA, 2025. Association for Computing Machinery. doi: 10.1145/3706598.3713705.

Ayelet Dassa. "Opening Our Time Capsule"—creating an individualized music and other memory cues database to promote communication between spouses and people with dementia during visits to a nursing home. *Frontiers in Medicine*, 5:215, 2018. doi: 10.3389/fmed.2018.00215.

Helén Dellkvist, Ana Luiza Dallora, Line Christiansen, and Lisa Skär. The use of a digital life story to support person-centred care of older adults with dementia: A scoping review. *Digital Health*, 10:20552076241241231, 2024. doi: 10.1177/20552076241241231.

Paul Denny, Hassan Khosravi, Arto Hellas, Juho Leinonen, and Sami Sarsa. Can we trust AI-generated educational content? Comparative analysis of human and AI-generated learning resources. arXiv preprint arXiv:2306.10509, 2023. URL `https://arxiv.org/abs/2306.10509`.

Design Council. What is the framework for innovation? Design Council's evolved Double Diamond. `https://www.designcouncil.org.uk/resources/framework-for-innovation/`, 2021. Accessed: 2026-08-03. 2021 version archived at `https://web.archive.org/web/2021/https://www.designcouncil.org.uk/news-opinion/what-framework-innovation-design-councils-evolved-double-diamond`.

Jan Dewing. Personhood and dementia: revisiting Tom Kitwood's ideas. *International Journal of Older People Nursing*, 3(1):3–13, 2008. doi: 10.1111/j.1748-3743.2007.00103.x.

Abhinav Dhall, Amanjot Kaur, Roland Goecke, and Tom Gedeon. EmotiW 2018: Audio-video, student engagement and group-level affect prediction. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, ICMI '18, pages 653–656, New York, NY, USA, 2018. Association for Computing Machinery. doi: 10.1145/3242969.3264993. arXiv preprint arXiv:1808.07773.

Cheryl Diermyer and Chris Blakesley. Story-based teaching and learning: Practices and technologies. In *Proceedings of the 25th Annual Conference on Distance Teaching & Learning*, Madison, WI, 2009. Board of Regents of the University of Wisconsin System.

Griffin Dietz, Jimmy K Le, Nadin Tamer, Jenny Han, Hyowon Gweon, Elizabeth L Murnane, and James A. Landay. Storycoder: Teaching computational thinking concepts through storytelling in a voice-guided app for children. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, page 54, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. doi: 10.1145/3411764.3445039.

Griffin Dietz, Nadin Tamer, Carina Ly, Jimmy K Le, and James A. Landay. Visual StoryCoder: A multimodal programming environment for children's creation of stories. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, page 96, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450394215. doi: 10.1145/3544548.3580981.

Betsy DiSalvo and Carl DiSalvo. Designing for democracy in education: Participatory design and the learning sciences. In *Learning and Becoming in Practice: The International Conference of the Learning Sciences (ICLS) 2014*, volume 2, pages 793–799, Boulder, CO, 2014. International Society of the Learning Sciences.

Kevin Doherty and Gavin Doherty. Engagement in HCI: Conception, theory and measurement. *ACM Computing Surveys*, 51(5):1–39, 2018. doi: 10.1145/3234149.

Dietrich Dörner and Joachim Funke. Complex problem solving: What it is and what it is not. *Frontiers in Psychology*, 8:1153, 2017. doi: 10.3389/fpsyg.2017.01153.

Stefania Druga, Sarah T. Vu, Eesh Likhith, and Tammy Qiu. Inclusive AI literacy for kids around the world. In *Proceedings of FabLearn 2019*, FL2019, pages 104–111, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450362443. doi: 10.1145/3311890.3311904.

Stefania Druga, Fee Lia Christoph, and Amy J Ko. Family as a third space for AI literacies: How do children and parents learn about AI together? In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, page 225, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391573. doi: 10.1145/3491102.3502031.

Stefania Druga, Jason Yip, Michael Preston, and Devin Dillon. The 4 As: Ask, Adapt, Author, Analyze: AI literacy framework for families. In Mizuko Ito, Remy Cross, Karthik Dinakar, and Candice Odgers, editors, *Algorithmic Rights and Protections for Children*, pages 193–232. The MIT Press, Cambridge, MA, 2023. doi: 10.7551/mitpress/13654.003.0014.

Mandi Dupain and Loréal L. Maguire. Health digital storytelling projects. *American Journal of Health Education*, 38(1):41–43, 2007. doi: 10.1080/19325037.2007.10598941.

Suzanne M. Dyer, Stephanie L. Harrison, Kate Laver, Craig Whitehead, and Maria Crotty. An overview of systematic reviews of pharmacological and non-pharmacological interventions for the treatment of behavioral and psychological symptoms of dementia. *International Psychogeriatrics*, 30(3):295–309, 2018. doi: 10.1017/S1041610217002344.

David Edvardsson, Deirdre Fetherstonhaugh, Rhonda Nay, and Stephen Gibson. Development and initial testing of the Person-centered Care Assessment Tool (P-CAT). *International Psychogeriatrics*, 22(1):101–108, 2010. doi: 10.1017/S1041610209990688.

Emmeline Edwards, Coryse St Hillaire-Clarke, David W. Frankowski, Robert Finkelstein, Thomas Cheever, Wen G. Chen, Lisa Onken, Amy Poremba, Robert Riddle, Dana Schloesser, Caitlin E. Burgdorf, Nena Wells, Renée Fleming, and Francis S. Collins. NIH music-based intervention toolkit: Music-based interventions for brain disorders of aging. *Neurology*, 100(18):868–878, 2023. doi: 10.1212/WNL.0000000000206797.

Douglas C. Engelbart. Augmenting human intellect: A conceptual framework. Technical Report AFOSR-3223, Stanford Research Institute, Menlo Park, CA, 1962.

K. Anders Ericsson and Herbert A. Simon. Verbal reports as data. *Psychological Review*, 87(3):215–251, 1980. doi: 10.1037/0033-295X.87.3.215.

Enrique Estellés-Arolas and Fernando González-Ladrón-de Guevara. Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2):189–200, 2012. doi: 10.1177/0165551512437638.

Patricia Ewick and Susan S. Silbey. Subversive stories and hegemonic tales: Toward a sociology of narrative. *Law & Society Review*, 29(2):197–226, 1995. doi: 10.2307/3054010.

Regina E. Fabry. Narrative scaffolding. *Review of Philosophy and Psychology*, 14(4):1147–1167, 2023. doi: 10.1007/s13164-021-00595-w.

Valerie Farnsworth. Conceptualizing identity, learning and social justice in community-based learning. *Teaching and Teacher Education*, 26(7):1481–1489, 2010. doi: 10.1016/j.tate.2010.06.006.

Fathom Video. Fathom AI notetaker, 2024. URL `https://fathom.ai/`. Accessed: 2026-08-03. Previously at `https://fathom.video`, accessed 2024-09-13.

Sam Fazio, Douglas Pace, Janice Flinner, and Beth Kallmyer. The fundamentals of person-centered care for individuals with dementia. *The Gerontologist*, 58(suppl_1):S10–S19, 2018. doi: 10.1093/geront/gnx122.

Rebecca Ferguson. Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6):304–317, 2012. doi: 10.1504/IJTEL.2012.051816.

Fireflies.ai. Fireflies: AI meeting assistant, 2024. URL https://fireflies.ai. Accessed: 2024-09-13; verified live 2026-08-03.

Kristie Fisher, Scott Counts, and Aniket Kittur. Distributed sensemaking: Improving sensemaking by leveraging the efforts of previous users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 247–256, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450310154. doi: 10.1145/2207676.2207711. URL https://doi.org/10.1145/2207676.2207711.

Hannah Flamm. "They Want Docile": How nursing homes in the United States overmedicate people with dementia. Technical report, Human Rights Watch, New York, February 2018. URL https://www.hrw.org/report/2018/02/05/they-want-docile/how-nursing-homes-in-the-united-states-overmedicate-people-with. Accessed 2026-08-03.

Bent Flyvbjerg. Five misunderstandings about case-study research. *Qualitative Inquiry*, 12 (2):219–245, 2006. doi: 10.1177/1077800405284363.

Catherine Twomey Fosnot, editor. *Constructivism: Theory, Perspectives, and Practice*. Teachers College Press, New York, second edition, 2013. ISBN 9780807772591. Ebook edition; second edition first published in print 2005.

Jane Fossey, Clive Ballard, Edmund Juszczak, Ian James, Nicola Alder, Robin Jacoby, and Robert Howard. Effect of enhanced psychosocial care on antipsychotic use in nursing home residents with severe dementia: cluster randomised trial. *BMJ*, 332(7544):756–761, 2006. doi: 10.1136/bmj.38782.575868.7c.

Jennifer A Fredricks, Phyllis C Blumenfeld, and Alison H Paris. School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1): 59–109, 2004. doi: 10.3102/00346543074001059.

Peter A. Frensch and Joachim Funke. Definitions, traditions, and a general framework for understanding complex problem solving. In Peter A. Frensch and Joachim Funke, editors, *Complex Problem Solving: The European Perspective*, pages 3–25. Lawrence Erlbaum Associates, Hillsdale, NJ, 1995.

Ashvin Gandhi, Huizi Yu, and David C Grabowski. High nursing staff turnover in nursing homes offers important quality information. *Health Affairs*, 40(3):384–391, 2021. doi: 10.1377/hlthaff.2020.00957.

Matthias Ganzinger, Nicola Kunz, Pascal Fuchs, Cornelia K Lyu, Martin Loos, Martin Dugas, and Thomas M Pausch. Automated generation of discharge summaries: leveraging large language models with clinical data. *Scientific Reports*, 15(1):16466, 2025. doi: 10.1038/s41598-025-01618-7.

Joseph E Gaugler and Lauren L Mitchell. Reimagining family involvement in residential long-term care. *Journal of the American Medical Directors Association*, 23(2):235–240, 2022. doi: 10.1016/j.jamda.2021.12.022.

F. Xavier Gaya-Morey, Jose M. Buades-Rubio, Philippe Palanque, Raquel Lacuesta, and Cristina Manresa-Yee. Deep learning-based facial expression recognition for the elderly: A systematic review. *Expert Systems with Applications*, 326:132709, 2026. doi: 10.1016/j.eswa.2026.132709. arXiv:2502.02618.

Nahum Gershon and Ward Page. What storytelling can do for information visualization. *Communications of the ACM*, 44(8):31–37, 2001. doi: 10.1145/381641.381653.

Stefania Giannini. Generative AI and the future of education. UNESCO, July 2023. URL https://unesdoc.unesco.org/ark:/48223/pf0000385877. Accessed: 2026-08-03.

Patricia R Gilman. Complex adaptive systems: A framework for an integrated chronic care model. *Advances in Nursing Science*, 44(4):330–339, 2021. doi: 10.1097/ANS.0000000000000380.

Manuela Glaser, Bärbel Garsoffky, and Stephan Schwan. Narrative-based learning: Possible benefits and problems. *Communications*, 34(4):429–447, 2009. doi: 10.1515/COMM.2009.026.

Doreen Goerss, Stefanie Köhler, Eleonora Rong, Anna Gesine Temp, Ingo Kilimann, Gerald Bieber, and Stefan Teipel. Smartwatch-based interventions for people with dementia: User-centered design approach. *JMIR Aging*, 7:e50107, 2024. doi: 10.2196/50107.

Cleotilde Gonzalez, Kate Donahue, Daniel G. Goldstein, Hoda Heidari, Mohammad S. Jalali, Beau Schelble, Aarti Singh, and Anita Williams Woolley. Toward a science of human–AI teaming for decision making: A complementarity framework. *PNAS Nexus*, 5(3):pgag030, 2026. doi: 10.1093/pnasnexus/pgag030.

Google. Women who choose Computer Science—What Really Matters: The critical role of encouragement and exposure, May 2014. URL https://static.googleusercontent.com/media/edu.google.com/en//pdfs/women-who-choose-what-really.pdf. Accessed: 2026-08-03.

Google Inc. and Gallup Inc. Diversity gaps in computer science: Exploring the underrepresentation of girls, Blacks and Hispanics, 2016. URL https://services.google.com/fh/files/misc/diversity-gaps-in-computer-science-report.pdf. Accessed: 2026-08-03.

Charles R. Graham. Theoretical considerations for understanding technological pedagogical content knowledge (TPACK). *Computers & Education*, 57(3):1953–1960, 2011. doi: 10.1016/j.compedu.2011.04.010.

Trisha Greenhalgh and Brian Hurwitz. Why study narrative? *BMJ*, 318(7175):48–50, 1999. doi: 10.1136/bmj.318.7175.48.

Sergio Grueso and Raquel Viejo-Sobera. Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: a systematic review. *Alzheimer's Research & Therapy*, 13(1):162, 2021. doi: 10.1186/s13195-021-00900-w.

Ken Gu, Madeleine Grunde-McLaughlin, Andrew McNutt, Jeffrey Heer, and Tim Althoff. How do data analysts respond to AI assistance? A Wizard-of-Oz study. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery. doi: 10.1145/3613904.3641891.

Fang Gui, Jiaoyun Yang, Qilin Wu, Yang Liu, Jia Zhou, and Ning An. Enhancing caregiver empowerment through the Story Mosaic system: Human-centered design approach for visualizing older adult life stories. *JMIR Aging*, 6:e50037, 2023. doi: 10.2196/50037.

Fang Gui, Jiaoyun Yang, Yiming Tang, Hongtu Chen, and Ning An. Structured life narratives: Building life story hierarchies with graph-enhanced event feature refinement. *Applied Sciences*, 14(2):918, 2024. doi: 10.3390/app14020918.

Philip J. Guo, Julia M. Markel, and Xiong Zhang. Learnersourcing at scale to overcome expert blind spots for introductory programming: A three-year deployment study on the Python Tutor website. In *Proceedings of the Seventh ACM Conference on Learning @ Scale*, pages 301–304, Virtual Event, USA, 2020. ACM. doi: 10.1145/3386527.3406733.

Abhay Gupta, Arjun D'Cunha, Kamal Awasthi, and Vineeth Balasubramanian. DAiSEE: Towards user engagement recognition in the wild, 2016. URL `https://arxiv.org/abs/1609.01885`. arXiv preprint, revised v7 July 2022; version 1 (2016) circulated as "DAISEE: Dataset for Affective States in E-Learning Environments" with authors Gupta, Jaiswal, Adhikari, and Balasubramanian.

Mark Guzdial. *Learner-Centered Design of Computing Education: Research on Computing for Everyone*. Synthesis Lectures on Human-Centered Informatics. Morgan & Claypool Publishers, San Rafael, CA, 2015. ISBN 9781627053518. doi: 10.2200/S00684ED1V01Y201511HCI033.

Emily Haesler, Michael Bauer, and Rhonda Nay. Staff–family relationships in the care of older people: a report on a systematic review. *Research in Nursing & Health*, 30(4):385–398, 2007. doi: 10.1002/nur.20200.

Jeffrey T. Hancock, Mor Naaman, and Karen Levy. AI-mediated communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication*, 25(1):89–100, 2020. doi: 10.1093/jcmc/zmz022.

Ben Hannigan and Margaret McAllister. The "wicked" global problems facing nursing. In Margaret McAllister and Donna Lee Brien, editors, *Empowerment Strategies for Nurses: Developing Resiliency in Practice*, pages 43–57. Springer Publishing Company, New York, second edition, 2019. ISBN 9780826167897. doi: 10.1891/9780826167927.0003.

Didik Hariyanto, Sigit Yatmono, Moh Khairudin, and Thomas Köhler. Students e-learning readiness towards education 4.0: Instrument development and validation. *Jurnal Pendidikan Vokasi*, 12(3):236–244, 2022. doi: 10.21831/jpv.v12i3.51798.

Sandra G. Hart and Lowell E. Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Peter A. Hancock and Najmedin Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139–183. North-Holland, Amsterdam, 1988. ISBN 9780444703880. doi: 10.1016/S0166-4115(08)62386-9.

Gillian R. Hayes. The relationship of action research to human-computer interaction. *ACM Transactions on Computer-Human Interaction*, 18(3):15:1–15:20, 2011. doi: 10.1145/1993060.1993065.

Janine K Hayward, Charlotte Gould, Emma Palluotto, Emily Kitson, Emily R Fisher, and Aimee Spector. Interventions promoting family involvement with care homes following placement of a relative with dementia: A systematic review. *Dementia*, 21(2):618–647, 2022. doi: 10.1177/14713012211046595.

Janine K Hayward, Charlotte Gould, Emma Palluotto, Emily C Kitson, and Aimee Spector. Family involvement with care homes following placement of a relative living with dementia: a review. *Ageing & Society*, 43(7):1530–1575, 2023. doi: 10.1017/S0144686X21000957.

Cheryl Heiks and Nicole Sabine. Long term care and skilled nursing facilities. *Delaware Journal of Public Health*, 8(5):144–149, 2022. doi: 10.32481/djph.2022.12.032.

Eric B. Hekler, Susan Michie, Misha Pavel, Daniel E. Rivera, Linda M. Collins, Holly B. Jimison, Claire Garnett, Skye Parral, and Donna Spruijt-Metz. Advancing models and theories for digital behavior change interventions. *American Journal of Preventive Medicine*, 51(5):825–832, 2016. doi: 10.1016/j.amepre.2016.06.013.

Julia Hermann, Aiden Danny Mäder, Ann-Kathrin Kubullek, Celine-Chiara Hey, and Aysegül Dogangün. (Intelligent) technical systems in elderly care: The caregivers perspective. In *Proceedings of the 34th Australian Conference on Human-Computer Interaction*, pages 81–87, New York, NY, USA, 2022. Association for Computing Machinery. doi: 10.1145/3572921.3572948.

Suzanne Hidi and K. Ann Renninger. The four-phase model of interest development. *Educational Psychologist*, 41(2):111–127, 2006. doi: 10.1207/s15326985ep4102_4.

Nicole T. M. Hill, Loren Mowszowski, Sharon L. Naismith, Verity L. Chadwick, Michael Valenzuela, and Amit Lampit. Computerized cognitive training in older adults with mild cognitive impairment or dementia: A systematic review and meta-analysis. *American Journal of Psychiatry*, 174(4):329–340, 2017. doi: 10.1176/appi.ajp.2016.16030360.

John H. Holland. Complex adaptive systems. *Daedalus*, 121(1):17–30, 1992. JSTOR: `https://www.jstor.org/stable/20025416`.

Kat Holmes. *Mismatch: How Inclusion Shapes Design*. Simplicity: Design, Technology, Business, Life. MIT Press, Cambridge, MA, 2020. ISBN 978-0-262-53948-7. Paperback edition; originally published 2018.

Kristina Höök and Jonas Löwgren. Strong concepts: Intermediate-level knowledge in interaction design research. *ACM Transactions on Computer-Human Interaction*, 19(3): 1–18, 2012. doi: 10.1145/2362364.2362371.

Min Hu and Hao Li. Student engagement in online learning: A review. In *2017 International Symposium on Educational Technology (ISET)*, pages 39–43, Hong Kong, 2017. IEEE. doi: 10.1109/ISET.2017.17.

Oliver Huang, Muhammad Fatir, Tianyu Luo, Sangho Suh, Hariharan Subramonyam, and Carolina Nobre. Narrative scaffolding: A narrative-first framework for data-driven sensemaking. In *Proceedings of the 31st International Conference on Intelligent User Interfaces*, IUI '26, pages 1112–1140, Paphos, Cyprus, 2026. ACM. doi: 10.1145/3742413.3789062.

Scott E. Hudson and Ian Smith. Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems. In *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work*, CSCW '96, pages 248–257. ACM, 1996. doi: 10.1145/240080.240295.

Min-Ling Hung, Chien Chou, Chao-Hsiu Chen, and Zang-Yuan Own. Learner readiness for online learning: Scale development and student perceptions. *Computers & Education*, 55(3):1080–1090, 2010. doi: 10.1016/j.compedu.2010.05.004.

Susanne Hupfer. Talent and workforce effects in the age of AI: Insights from Deloitte's State of AI in the Enterprise, 2nd edition survey. Technical report, Deloitte Insights, Deloitte Center for Technology, Media & Telecommunications, March 2020. URL `https://www2.deloitte.com/content/dam/insights/us/articles/6546_talent-and-workforce-effects-in-the-age-of-ai/DI_Talent-and-workforce-effects-in-the-age-of-AI.pdf`. Accessed 2026-08-03.

Matti Hyvärinen. Prototypes, genres, and concepts: Travelling with narratives. *Narrative Works*, 2(1):10–32, 2012. URL `https://journals.lib.unb.ca/index.php/NW/article/view/19496`.

Anna Sigridur Islind, Helena Vallo Hult, Victoria Johansson, Eva Angenete, and Martin Gellerstedt. Invisible work meets visible work: Infrastructuring from the perspective of patients and healthcare professionals. In *Proceedings of the 54th Hawaii International Conference on System Sciences*, pages 3556–3565. Hawaii International Conference on System Sciences, 2021. doi: 10.24251/hicss.2021.431.

Mizuko Ito, Kris Gutiérrez, Sonia Livingstone, Bill Penuel, Jean Rhodes, Katie Salen, Juliet Schor, Julian Sefton-Green, and S. Craig Watkins. Connected learning: An agenda for research and design. Technical report, Digital Media and Learning Research Hub, Irvine, CA, 2013.

Jörn-Henrik Jacobsen, Johannes Stelzer, Thomas Hans Fritz, Gael Chételat, Renaud La Joie, and Robert Turner. Why musical memory can be preserved in advanced Alzheimer's disease. *Brain*, 138(8):2438–2450, 2015. doi: 10.1093/brain/awv135.

Michael J. Jacobson, James A. Levin, and Manu Kapur. Education as a complex system: Conceptual and methodological implications. *Educational Researcher*, 48(2):112–119, 2019. doi: 10.3102/0013189X19826958.

P. Devereaux Jennings and Royston Greenwood. Constructing the iron cage: Institutional theory and enactment. In Robert I. Westwood and Stewart R. Clegg, editors, *Debating Organization: Point-Counterpoint in Organization Studies*, pages 195–207. Blackwell, Malden, MA, 2003. ISBN 978-0631216933.

Hawoong Jeong, Zoltán Néda, and Albert-László Barabási. Measuring preferential attachment in evolving networks. *Europhysics Letters*, 61(4):567–572, 2003. doi: 10.1209/epl/i2003-00166-9.

Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38, 2023. doi: 10.1145/3571730.

Di Jiang, Luowei Yan, and Florence Mayrand. Emotion expressions and cognitive impairments in the elderly: review of the contactless detection approach. *Frontiers in Digital Health*, 6:1335289, 2024. doi: 10.3389/fdgth.2024.1335289.

Martin Johnson and Dominika Majewska. Formal, non-formal, and informal learning: What are they, and how can we research them? Research report, Cambridge University Press & Assessment, Cambridge, UK, September 2022. URL `https://www.cambridgeassessment.org.uk/Images/665425-formal-non-formal-and-informal-learning-what-are-they-and-how-can-we-research-them-.pdf`. ERIC Number: ED626005.

Cindy Jones, Billy Sung, and Wendy Moyle. Assessing engagement in people with dementia: a new approach to assessment using video analysis. *Archives of Psychiatric Nursing*, 29(6):377–382, 2015. doi: 10.1016/j.apnu.2015.06.019.

Parunyou Julayanont and Ziad S. Nasreddine. Montreal Cognitive Assessment (MoCA): Concept and clinical review. In A. J. Larner, editor, *Cognitive Screening Instruments: A Practical Approach*, pages 139–195. Springer, Cham, 2nd edition, 2017. doi: 10.1007/978-3-319-44775-9_7.

Carolin Kaiser, Jakob Kaiser, René Schallner, and Sabrina Schneider. A new era of online search? a large-scale study of user behavior and personal preferences during practical search tasks with generative AI versus traditional search engines. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '25, New York, NY, 2025. ACM. doi: 10.1145/3706599.3720123.

Rie Kamikubo, Lining Wang, Crystal Marte, Amnah Mahmood, and Hernisa Kacorri. Data representativeness in accessibility datasets: A meta-analysis. In *Proceedings of the 24th*

International ACM SIGACCESS Conference on Computers and Accessibility, pages 1–15, Athens, Greece, 2022. Association for Computing Machinery. doi: 10.1145/3517428.3544826.

Harmanpreet Kaur, Eytan Adar, Eric Gilbert, and Cliff Lampe. Sensible AI: Re-imagining interpretability and explainability using sensemaking theory. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, pages 702–714. Association for Computing Machinery, 2022. doi: 10.1145/3531146.3533135.

Shehroz S. Khan, Tong Zhu, Bing Ye, Alex Mihailidis, Andrea Iaboni, Kristine Newman, Angel He Wang, and Lori Schindel Martin. DAAD: A framework for detecting agitation and aggression in people living with dementia using a novel multi-modal sensor network. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 703–710. IEEE, 2017. doi: 10.1109/ICDMW.2017.98.

Hassan Khosravi, Gianluca Demartini, Shazia Sadiq, and Dragan Gasevic. Charting the design and analytics agenda of learnersourcing systems. In *LAK21: 11th International Learning Analytics and Knowledge Conference*, pages 32–42, Irvine, CA, USA, 2021. Association for Computing Machinery. doi: 10.1145/3448139.3448143.

Hassan Khosravi, Paul Denny, Steven Moore, and John Stamper. Learnersourcing in the age of AI: Student, educator and machine partnerships for content creation. *Computers and Education: Artificial Intelligence*, 5:100151, 2023. doi: 10.1016/j.caeai.2023.100151.

Aman Khullar, Nikhil Nalin, Abhishek Prasad, Ann John Mampilli, and Neha Kumar. Nurturing capabilities: Unpacking the gap in human-centered evaluations of AI-based systems. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–18, Yokohama, Japan, 2025. Association for Computing Machinery. doi: 10.1145/3706598.3713278.

Juho Kim. *Learnersourcing: Improving Learning with Collective Learner Activity*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 2015. URL `https://hdl.handle.net/1721.1/101464`.

Jacqueline Kindell, Simon Burrow, Ray Wilkinson, and John David Keady. Life story resources in dementia care: a review. *Quality in Ageing and Older Adults*, 15(3):151–161, 2014. doi: 10.1108/QAOA-02-2014-0003.

Aniket Kittur, Boris Smus, Susheel Khamkar, and Robert E. Kraut. CrowdForge: Crowdsourcing complex work. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, pages 43–52, New York, NY, USA, 2011. ACM. doi: 10.1145/2047196.2047202.

Tom Kitwood. Brain, mind and dementia: with particular reference to Alzheimer's disease. *Ageing & Society*, 9(1):1–15, 1989. doi: 10.1017/S0144686X00013337.

Tom Kitwood. *Dementia Reconsidered: The Person Comes First*. Open University Press, Buckingham, UK, 1997. ISBN 978-0-335-19855-9.

Maria Kjærup, Mikael B. Skov, Peter Axel Nielsen, Jesper Kjeldskov, Jens Gerken, and Harald Reiterer. Longitudinal studies in HCI research: A review of CHI publications from 1982–2019. In *Advances in Longitudinal HCI Research*, pages 11–39. Springer International Publishing, 2021. doi: 10.1007/978-3-030-67322-2_2.

Gary Klein. Naturalistic decision making. *Human Factors*, 50(3):456–460, 2008. doi: 10.1518/001872008X288385.

Gary Klein, Jennifer K. Phillips, Erica L. Rall, and Deborah A. Peluso. A data-frame theory of sensemaking. In Robert R. Hoffman, editor, *Expertise Out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making*, pages 113–155. Lawrence Erlbaum Associates, 2007.

Matthew J. Koehler and Punya Mishra. What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education*, 9(1):60–70, 2009.

Dimitra Kokotsaki, Victoria Menzies, and Andy Wiggins. Project-based learning: A review of the literature. *Improving Schools*, 19(3):267–277, 2016. doi: 10.1177/1365480216659733.

Ann Kolanowski, Marie Boltz, Elizabeth Galik, Laura N. Gitlin, Helen C. Kales, Barbara Resnick, Kimberly S. Van Haitsma, Amy Knehans, Jane E. Sutterlin, Justine S. Sefcik, Wen Liu, Darina V. Petrovsky, Lauren Massimo, Andrea Gilmore-Bykovskyi, Margaret MacAndrew, Glenna Brewster, Vycki Nalls, Ying-Ling Jao, Naomi Duffort, and Danny Scerpella. Determinants of behavioral and psychological symptoms of dementia: A scoping review of the evidence. *Nursing Outlook*, 65(5):515–529, 2017. doi: 10.1016/j.outlook.2017.06.006.

David A. Kolb. *Experiential Learning: Experience as the Source of Learning and Development*. Pearson FT Press, Upper Saddle River, NJ, second edition, 2014. ISBN 978-0-13-389240-6. First edition published 1984 by Prentice-Hall.

Maria Kordaki and Panagiotis Kakavas. Digital storytelling as an effective framework for the development of computational thinking skills. In *EDULEARN17 Proceedings*, pages 6325–6335. IATED, 2017. doi: 10.21125/edulearn.2017.2435.

Torsten Kratz. The diagnosis and treatment of behavioral disorders in dementia. *Deutsches Ärzteblatt International*, 114(26):447–454, 2017. doi: 10.3238/arztebl.2017.0447.

Gautam Kunapuli. *Ensemble Methods for Machine Learning*. Manning Publications, 2023. ISBN 978-1-61729-713-7.

Karan S. Kverno, Betty S. Black, Marie T. Nolan, and Peter V. Rabins. Research on treating neuropsychiatric symptoms of advanced dementia with non-pharmacological strategies, 1998–2008: a systematic literature review. *International Psychogeriatrics*, 21(5):825–843, 2009. doi: 10.1017/S1041610209990196.

Labellerr. Labellerr: AI-Powered Data Labeling Platform, 2024. URL `https://www.labellerr.com/`. Accessed: 2024-10-30.

Nouhaila Labzour, Sanaa El Fkihi, Soukayna Benaissa, Yahya Zennayi, and Omar Bourja. A survey on facial emotion recognition for the elderly. In Saad Motahhir and Badre Bossoufi, editors, *Digital Technologies and Applications: Proceedings of ICDTA'23, Fez, Morocco, Volume 1*, volume 668 of *Lecture Notes in Networks and Systems*, pages 561–575, Cham, 2023. Springer. doi: 10.1007/978-3-031-29857-8_57.

Claudia K. Y. Lai, Ayumi Igarashi, Clare T. K. Yu, and Kenny C. W. Chin. Does life story work improve psychosocial well-being for older adults in the community? A quasi-experimental study. *BMC Geriatrics*, 18(1):119, 2018. doi: 10.1186/s12877-018-0797-0.

R. Eric Landrum, Karen Brakke, and Maureen A. McCarthy. The pedagogical power of storytelling. *Scholarship of Teaching and Learning in Psychology*, 5(3):247–253, 2019. doi: 10.1037/stl0000152.

Alyssa M. Lanzi, Anna K. Saylor, Davida Fromm, Houjun Liu, Brian MacWhinney, and Matthew L. Cohen. DementiaBank: Theoretical rationale, protocol, and illustrative analyses. *American Journal of Speech-Language Pathology*, 32(2):426–438, 2023. doi: 10.1044/2022_AJSLP-22-00281.

Jean Lave and Etienne Wenger. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press, Cambridge, UK, 1991. doi: 10.1017/CBO9780511815355.

M. Powell Lawton, Kimberly Van Haitsma, Margaret A. Perkinson, and Katy Ruckdeschel. Observed affect and quality of life in dementia: Further affirmations and problems. *Journal of Mental Health and Aging*, 5(1):69–81, 1999.

Christopher A. Le Dantec and Carl DiSalvo. Infrastructuring and the formation of publics in participatory design. *Social Studies of Science*, 43(2):241–264, 2013. doi: 10.1177/0306312712471581.

Christopher A. Le Dantec and Sarah Fox. Strangers at the gate: Gaining access, building rapport, and co-constructing community-based research. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, pages 1348–1358, New York, NY, USA, 2015. Association for Computing Machinery. doi: 10.1145/2675133.2675147.

Sarah Lebovitz, Natalia Levina, and Hila Lifshitz-Assaf. Is AI ground truth really true? The dangers of training and evaluating AI tools based on experts' know-what. *MIS Quarterly*, 45(3):1501–1526, 2021. doi: 10.25300/MISQ/2021/16564.

Chan Jean Lee. The test taker's fallacy: How students guess answers on multiple-choice tests. *Journal of Behavioral Decision Making*, 32(2):140–151, 2019. doi: 10.1002/bdm.2101.

Hao-Ping Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, 2025. doi: 10.1145/3706598.3713778.

Raymond S. T. Lee, James N. K. Liu, Karo S. Y. Yeung, Alan H. L. Sin, and Dennis T. F. Shum. Agent-based web content engagement time (WCET) analyzer on e-publication system. In *2009 Ninth International Conference on Intelligent Systems Design and Applications*, pages 67–72, Pisa, Italy, 2009. IEEE. doi: 10.1109/ISDA.2009.189.

Florian Lehmann, Krystsina Shauchenka, and Daniel Buschek. Collaborative document editing with multiple users and AI agents. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*. ACM, 2026. doi: 10.1145/3772318.3790648.

Justin B. Leibowitz, Charity Flener Lovitt, and Craig S. Seager. Development and validation of a survey to assess belonging, academic engagement, and self-efficacy in STEM RLCs. *Learning Communities: Research & Practice*, 8(1):3, 2020. ERIC Number: EJ1251571.

Naomi Ehrich Leonard, Anastasia Bizyaeva, and Alessio Franci. Fast and flexible multiagent decision-making. *Annual Review of Control, Robotics, and Autonomous Systems*, 7 (1):19–45, 2024. doi: 10.1146/annurev-control-090523-100059.

James C. Lester, Hiller A. Spires, John L. Nietfeld, James Minogue, Bradford W. Mott, and Eleni V. Lobene. Designing game-based learning environments for elementary science education: A narrative-centered learning perspective. *Information Sciences*, 264:4–18, 2014. doi: 10.1016/j.ins.2013.09.005.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems*, volume 33, pages 9459–9474. Curran Associates, Inc., 2020.

Mariana Leyton Escobar, Piet A. M. Kommers, and Ardion Beldad. Using narratives as tools for channeling participation in online communities. *Computers in Human Behavior*, 37:64–72, 2014. doi: 10.1016/j.chb.2014.04.013.

Hui-Chi Li, Hsiu-Hung Wang, Chu-Yun Lu, Tai-Been Chen, Yu-Hua Lin, and I Lee. The effect of music therapy on reducing depression in people with dementia: A systematic review and meta-analysis. *Geriatric Nursing*, 40(5):510–516, 2019. doi: 10.1016/j.gerinurse.2019.03.017.

Charlotte Linde. Narrative and social tacit knowledge. *Journal of Knowledge Management*, 5(2):160–171, 2001. doi: 10.1108/13673270110393202.

David A Lindeman, Katherine K Kim, Caroline Gladstone, and Ester Carolina Apesoa-Varano. Technology and caregiving: Emerging interventions and directions for research. *The Gerontologist*, 60(Supplement_1):S41–S49, 2020. doi: 10.1093/geront/gnz178.

Alex Lishinski and Joshua Rosenberg. All the pieces matter: The relationship of momentary self-efficacy and affective experiences with CS1 achievement and interest in computing. In *Proceedings of the 17th ACM Conference on International Computing Education Research*, pages 252–265, New York, NY, USA, 2021. Association for Computing Machinery. doi: 10.1145/3446871.3469740.

Breanne K. Litts, Kristin A. Searle, Bryan M. J. Brayboy, and Yasmin B. Kafai. Computing for all?: Examining critical biases in computational tools for learning. *British Journal of Educational Technology*, 52(2):842–857, 2021. doi: 10.1111/bjet.13059.

Michael Xieyang Liu, Tongshuang Wu, Tianying Chen, Franklin Mingzhe Li, Aniket Kittur, and Brad A. Myers. Selenite: Scaffolding online sensemaking with comprehensive overviews elicited from large language models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM, 2024. doi: 10.1145/3613904.3642149.

Genevieve Liveley. *Narratology*. Oxford University Press, Oxford, 2019. doi: 10.1093/oso/9780199687701.001.0001.

Chung Kwan Lo. What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4):410, 2023. doi: 10.3390/educsci13040410.

Duri Long and Brian Magerko. What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–16, New York, NY, USA, 2020. Association for Computing Machinery. doi: 10.1145/3313831.3376727.

Duri Long, Takeria Blunt, and Brian Magerko. Co-designing AI literacy exhibits for informal learning spaces. *Proceedings of the ACM on Human-Computer Interaction*, 5 (CSCW2):1–35, 2021. doi: 10.1145/3476034.

Tao Long, Sitong Wang, Julien Porquet, Yue Jiang, Jason Wu, Tuhin Chakrabarty, Savvas Petridis, David Ledo, Dingzeyu Li, Tiffany Tseng, Jeffrey V. Nickerson, and Lydia B. Chilton. Facilitating longitudinal research in HCI. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM, 2025. Workshop.

Xiaoli Lu. Coping with uncertainty: Towards an institutional sensemaking model. In *Managing Uncertainty in Crisis*, pages 13–34. Springer Singapore, Singapore, 2017. doi: 10.1007/978-981-10-3990-4_2.

Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. MediaPipe: A framework for building perception pipelines, 2019. arXiv preprint arXiv:1906.08172.

Dirk R. M. Lukkien, Sima Ipakchian Askari, Nathalie E. Stolwijk, Bob M. Hofstede, Henk Herman Nap, Wouter P. C. Boon, Alexander Peine, Ellen H. M. Moors, and Mirella M. N. Minkman. Making co-design more responsible: Case study on the development of an AI-based decision support system in dementia care. *JMIR Human Factors*, 11: e55961, 2024. doi: 10.2196/55961.

Yao Lyu, Jie Cai, and John M. Carroll. A systematic literature review of infrastructure studies in SIGCHI. *Proceedings of the ACM on Human-Computer Interaction*, 9(7): 1–41, 2025. doi: 10.1145/3757545.

Fiona Macleod, Lesley Storey, Teresa Rushe, and Katrina McLaughlin. Towards an increased understanding of reminiscence therapy for people with dementia: A narrative analysis. *Dementia*, 20(4):1375–1407, 2021. doi: 10.1177/1471301220941275.

Khalid Mahmood. Do people overestimate their information literacy skills? A systematic review of empirical evidence on the Dunning-Kruger effect. *Communications in Information Literacy*, 10(2):199–213, 2016. doi: 10.15760/comminfolit.2016.10.2.24.

Sally Maitlis and Marlys Christianson. Sensemaking in organizations: Taking stock and moving forward. *Academy of Management Annals*, 8(1):57–125, 2014. doi: 10.5465/19416520.2014.873177.

Jill Manthorpe and Kritika Samsi. Person-centered dementia care: current perspectives. *Clinical Interventions in Aging*, 11:1733–1740, 2016. doi: 10.2147/CIA.S104618.

James G. March and Herbert A. Simon. *Organizations*. Wiley, New York, 1958.

Lina Markauskaite, Rebecca Marrone, Oleksandra Poquet, Simon Knight, Roberto Martinez-Maldonado, Sarah Howard, Jo Tondeur, Maarten De Laat, Simon Buckingham Shum, Dragan Gašević, and George Siemens. Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI? *Computers and Education: Artificial Intelligence*, 3:100056, 2022. doi: 10.1016/j.caeai.2022.100056.

Luis Marquez-Carpintero, Sergio Suescun-Ferrandiz, Carolina Lorenzo Álvarez, Jorge Fernandez-Herrero, Diego Viejo, Rosabel Roig-Vila, and Miguel Cazorla. DIPSER: A dataset for in-person student engagement recognition in the wild, 2025. arXiv preprint arXiv:2502.20209.

Brendan Martin, Nicole Kaminski-Ozturk, Charlie O'Hara, and Richard Smiley. Examining the impact of the COVID-19 pandemic on burnout and stress among U.S. nurses. *Journal of Nursing Regulation*, 14(1):4–12, 2023. doi: 10.1016/S2155-8256(23)00063-7.

Allison Master, Sapna Cheryan, Adriana Moscatelli, and Andrew N. Meltzoff. Programming experience promotes higher STEM motivation among first-grade girls. *Journal of Experimental Child Psychology*, 160:92–106, 2017. doi: 10.1016/j.jecp.2017.03.013.

Anna Maria Matziorinis and Stefan Koelsch. The promise of music therapy for Alzheimer's disease: A review. *Annals of the New York Academy of Sciences*, 1516(1):11–17, 2022. doi: 10.1111/nyas.14864.

Orii McDermott, Nadia Crellin, Hanne Mette Ridder, and Martin Orrell. Music therapy in dementia: a narrative synthesis systematic review. *International Journal of Geriatric Psychiatry*, 28(8):781–794, 2013. doi: 10.1002/gps.3895.

Jane McKeown, Amanda Clarke, Christine Ingleton, Tony Ryan, and Julie Repper. The use of life story work with people with dementia to enhance person-centred care. *International Journal of Older People Nursing*, 5(2):148–158, 2010. doi: 10.1111/j.1748-3743.2010.00219.x.

Scott W. McQuiggan, Jonathan P. Rowe, Sunyoung Lee, and James C. Lester. Story-based learning: The impact of narrative on learning experiences and outcomes. In *Intelligent Tutoring Systems: 9th International Conference, ITS 2008, Montreal, Canada, June 23–27, 2008, Proceedings*, volume 5091 of *Lecture Notes in Computer Science*, pages 530–539, Berlin, Heidelberg, 2008. Springer. doi: 10.1007/978-3-540-69132-7_56.

Scott W. McQuiggan, Jennifer L. Robison, and James C. Lester. Affective transitions in narrative-centered learning environments. *Journal of Educational Technology & Society*, 13(1):40–53, 2010.

Donella H. Meadows. Leverage points: Places to intervene in a system. Technical report, The Sustainability Institute, Hartland, VT, 1999. URL `https://donellameadows.org/archives/leverage-points-places-to-intervene-in-a-system/`. A shorter version appeared as "Places to Intervene in a System," *Whole Earth*, no. 91 (Winter 1997), 78–84.

Donella H. Meadows. *Thinking in Systems: A Primer*. Chelsea Green Publishing, White River Junction, VT, 2008. Edited by Diana Wright.

Lingrui Mei, Jiayu Yao, Yuyao Ge, Yiwei Wang, Baolong Bi, Yujun Cai, Jiazhi Liu, Mingyu Li, Zhong-Zhi Li, Duzhen Zhang, Chenlin Zhou, Jiayi Mao, Tianze Xia, Jiafeng Guo, and Shenghua Liu. A survey of context engineering for large language models, 2025. arXiv preprint arXiv:2507.13334; not peer-reviewed.

Molly Millians. Learning readiness. In Sam Goldstein and Jack A. Naglieri, editors, *Encyclopedia of Child Behavior and Development*, pages 877–879. Springer, Boston, MA, 2011. doi: 10.1007/978-0-387-79061-9_1631.

Richard Milne and Carol Brayne. We need to think about data governance for dementia research in a digital era. *Alzheimer's Research & Therapy*, 12(1):17, 2020. doi: 10.1186/s13195-020-0584-y.

Joy Ming, Hawi H Tolera, Jiamin Tu, Ella Yitzhaki, Chit Sum Eunice Ngai, Madeline Sterling, Ariel C Avgar, Aditya Vashistha, and Nicola Dell. Exploring data-driven advocacy in home health care work. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–17, New York, NY, USA, 2025. Association for Computing Machinery. doi: 10.1145/3706598.3713086.

Melanie Mitchell. *Complexity: A guided tour*. Oxford University Press, 2009. doi: 10.1093/oso/9780195124415.001.0001.

Ali Mollahosseini, Behzad Hasani, and Mohammad H. Mahoor. AffectNet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 10(1):18–31, 2019. doi: 10.1109/TAFFC.2017.2740923.

Robert (Munro) Monarch. *Human-in-the-Loop Machine Learning: Active Learning and Annotation for Human-Centered AI*. Manning Publications, 2021. ISBN 978-1-61729-674-1.

Dylan Moore and Elizabeth Murnane. Self-portrait adaptive re-clustering (SPARC): An AI-centric framework for scaling collaborative meaning making in networked learning environments. In *Learnersourcing: Student-Generated Content @ Scale, Workshop at the Twelfth ACM Conference on Learning @ Scale (L@S '25)*, 2025.

Dylan Edward Moore, Sophia R. R. Moore, Bansharee Ireen, Winston P. Iskandar, Grigory Artazyan, and Elizabeth L. Murnane. Teaching artificial intelligence in extracurricular contexts through narrative-based learnersourcing. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, page 270, Honolulu, HI, USA, 2024. Association for Computing Machinery. doi: 10.1145/3613904.3642198.

Dylan Edward Moore, Songyun Tao, Sophia R. R. Moore, Brian Morgan, Dio Tadin, Christina Sapp Tadin, and Elizabeth L. Murnane. CareInsights: AI-enabled infrastructure for person-centered dementia care in resource-constrained facilities. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 9(4):202, 2025a. doi: 10.1145/3770687.

Dylan Edward Moore, Songyun Tao, Emma Ricci-De Lucca, Christina Sapp Tadin, Dio Tadin, Brian Morgan, and Elizabeth L. Murnane. Family In The Loop: Enabling family involvement and person-centered dementia care at long-term care facilities with collaborative AI tools. *Proceedings of the ACM on Human-Computer Interaction*, 9(7):1–45, 2025b. doi: 10.1145/3757599.

Celia Moreno-Morales, Raul Calero, Pedro Moreno-Morales, and Cristina Pintado. Music Therapy in the Treatment of Dementia: A Systematic Review and Meta-Analysis. *Frontiers in Medicine*, 7:160, 2020. ISSN 2296-858X. doi: 10.3389/fmed.2020.00160.

Meredith Ringel Morris, Michael S. Bernstein, Jeffrey P. Bigham, Amy S. Bruckman, and Andrés Monroy-Hernández. Is human-AI interaction CSCW? In *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing*, pages 95–97, New York, NY, 2024. ACM. doi: 10.1145/3678884.3689134.

Anna Moutafidou and Tharrenos Bratitsis. Digital storytelling: Giving voice to socially excluded people in various contexts. In *Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*, pages 219–226, New York, NY, USA, 2018. Association for Computing Machinery. doi: 10.1145/3218585.3218684.

Tamara Mujirishvili, Caterina Maidhof, Francisco Florez-Revuelta, Martina Ziefle, Miguel Richart-Martinez, and Julio Cabrero-García. Acceptance and privacy perceptions toward video-based active and assisted living technologies: Scoping review. *Journal of Medical Internet Research*, 25:e45297, 2023. doi: 10.2196/45297.

Kasia Muldner, Winslow Burleson, and Kurt VanLehn. "yes!": Using tutor and sensor data to predict moments of delight during instructional activities. In *User Modeling, Adaptation, and Personalization: 18th International Conference, UMAP 2010*, volume 6075 of *Lecture Notes in Computer Science*, pages 159–170, Big Island, HI, USA, 2010. Springer. doi: 10.1007/978-3-642-13470-8_16.

Mahin Naderifar, Hamideh Goli, and Fereshteh Ghaljaie. Snowball sampling: A purposeful method of sampling in qualitative research. *Strides in Development of Medical Education*, 14(3), 2017. doi: 10.5812/sdme.67670.

National Academies of Sciences, Engineering, and Medicine. *Human-AI Teaming: State-of-the-Art and Research Needs*. The National Academies Press, Washington, DC, 2022. ISBN 978-0-309-27017-5. doi: 10.17226/26355.

Davy Tsz Kit Ng, Jac Ka Lok Leung, Samuel Kai Wah Chu, and Maggie Shen Qiao. Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2:100041, 2021. doi: 10.1016/j.caeai.2021.100041.

Davy Tsz Kit Ng, Jac Ka Lok Leung, Maggie Jiahong Su, Iris Heung Yue Yim, Maggie Shen Qiao, and Samuel Kai Wah Chu. *AI Literacy in K-16 Classrooms*. Springer, Cham, Switzerland, 2022a. ISBN 978-3-031-18879-4. doi: 10.1007/978-3-031-18880-0.

Davy Tsz Kit Ng, Wanying Luo, Helen Man Yi Chan, and Samuel Kai Wah Chu. Using digital story writing as a pedagogy to develop AI literacy among primary students. *Computers and Education: Artificial Intelligence*, 3:100054, 2022b. doi: 10.1016/j.caeai.2022.100054.

Jakob Nielsen. *Usability Engineering*. Morgan Kaufmann, 1994. ISBN 978-0-12-518406-9. First published 1993 by Academic Press.

Hannele Niemi. Active learning—a cultural change needed in teacher education and schools. *Teaching and Teacher Education*, 18(7):763–780, 2002. doi: 10.1016/S0742-051X(02)00042-2.

Natalia Norori, Qiyang Hu, Florence Marcelle Aellen, Francesca Dalia Faraci, and Athina Tzovara. Addressing bias in big data and AI for health care: A call for open science. *Patterns*, 2(10):100347, 2021. doi: 10.1016/j.patter.2021.100347.

Wendy Norris, Amy Voida, and Stephen Voida. People talk in stories. Responders talk in data: A framework for temporal sensemaking in time- and safety-critical work. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1):108, April 2022. doi: 10.1145/3512955.

Shakked Noy and Whitney Zhang. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654):187–192, 2023. doi: 10.1126/science.adh2586.

Heather L O'Brien, Paul Cairns, and Mark Hall. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies*, 112:28–39, 2018. doi: 10.1016/j.ijhcs.2018.01.004.

Arnel B Ocay. Investigating the Dunning-Kruger effect among students within the contexts of a narrative-centered game-based learning environment. In *Proceedings of the 2019*

2nd International Conference on Education Technology Management, pages 8–13, New York, NY, USA, 2019. ACM. doi: 10.1145/3375900.3375918.

David L Olson and Kirsten Rosacker. Crowdsourcing and open source software participation. *Service Business*, 7(4):499–511, 2013. doi: 10.1007/s11628-012-0176-4.

Thomas A. O'Neill, Nathan J. McNeese, Amy Barron, and Beau Schelble. Human–autonomy teaming: A review and analysis of the empirical literature. *Human Factors*, 64(5):904–938, 2022. doi: 10.1177/0018720820960865.

Walter J. Ong. *Orality and Literacy: The Technologizing of the Word*. Methuen, London, 1982.

Julian E. Orr. *Talking About Machines: An Ethnography of a Modern Job*. Collection on Technology and Work. ILR Press, Ithaca, NY, 1996. ISBN 9780801483905.

Cassandra Overney, Hang Jiang, Urooj Haider, Cassandra Moe, Jasmine Mangat, Frank Pantano, Effie G. McMillian, Paul Riggins, and Nabeel Gillani. Human-AI narrative synthesis to foster shared understanding in civic decision-making. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*. ACM, 2026. doi: 10.1145/3772318.3791570.

Erin Palm, Astrit Manikantan, Herprit Mahal, Srikanth Subramanya Belwadi, and Mark E. Pepin. Assessing the quality of AI-generated clinical notes: Validated evaluation of a large language model ambient scribe. *Frontiers in Artificial Intelligence*, 8:1691499, 2025. doi: 10.3389/frai.2025.1691499. arXiv:2505.17047.

Zachary A. Pardos, Ioannis Anastasopoulos, and Shreya K. Sheel. Conducting rapid experimentation with an open-source adaptive tutoring system. In Ning Wang, Genaro Rebolledo-Mendez, Vania Dimitrova, Noboru Matsuda, and Olga C. Santos, editors, *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky (AIED 2023)*, volume 1831 of *Communications in Computer and Information Science*, pages 38–43, Cham, Switzerland, 2023. Springer. doi: 10.1007/978-3-031-3633 6-8_5.

Joon Sung Park, Michael S. Bernstein, Robin N. Brewer, Ece Kamar, and Meredith Ringel Morris. Understanding the representation and representativeness of age in ai data sets. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '21, page 834–842, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450384735. doi: 10.1145/3461702.3462590. URL `https://doi.org/10.1145/3461702.3462590`.

Joon Sung Park, Joseph C. O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*, pages 1–22, New York, NY, 2023. ACM. doi: 10.1145/358618 3.3606763.

Sharoda A. Paul and Meredith Ringel Morris. CoSense: Enhancing sensemaking for collaborative web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1771–1780, New York, NY, USA, 2009. Association for Computing Machinery. doi: 10.1145/1518701.1518974.

Sharoda A. Paul and Madhu C. Reddy. Understanding together: Sensemaking in collaborative information seeking. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, CSCW '10, pages 321–330. Association for Computing Machinery, 2010. doi: 10.1145/1718918.1718976.

Roy D. Pea. The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences*, 13(3):423–451, 2004. doi: 10.1207/s15327809jls1303_6.

E. Martin Pedersen. Storytelling and the art of teaching. *English Teaching Forum*, 33(1): 2–5, 1995.

Nancy E. Perry. Young children's self-regulated learning and contexts that support it. *Journal of Educational Psychology*, 90(4):715–729, 1998. doi: 10.1037/0022-0663.90.4.715.

Diane Pesco and Andréanne Gagné. Scaffolding narrative skills: A meta-analysis of instruction in early childhood settings. *Early Education and Development*, 28(7):773–793, 2017. doi: 10.1080/10409289.2015.1060800.

Paul R. Pintrich, David A. F. Smith, Teresa Garcia, and Wilbert J. McKeachie. A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ). Technical Report 91-B-004, National Center for Research to Improve Postsecondary Teaching and Learning, University of Michigan, Ann Arbor, MI, 1991. URL `https://eric.ed.gov/?id=ED338122`. ERIC Document ED338122. Accessed: 2026-08-03.

Volkmar Pipek and Volker Wulf. Infrastructuring: Toward an integrated perspective on the design and use of information technology. *Journal of the Association for Information Systems*, 10(5):447–473, 2009. doi: 10.17705/1jais.00195.

Peter Pirolli and Stuart Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of the 2005 International Conference on Intelligence Analysis*, McLean, VA, 2005.

Peter Pirolli and Daniel M. Russell. Introduction to this special issue on sensemaking. *Human-Computer Interaction*, 26(1):1–8, 2011. doi: 10.1080/07370024.2011.556557.

Nea Pirttinen, Paul Denny, Arto Hellas, and Juho Leinonen. Lessons learned from four computing education crowdsourcing systems. *IEEE Access*, 11:22982–22992, 2023. doi: 10.1109/ACCESS.2023.3253642.

Plato. *The Republic*. Clarendon Press, Oxford, 1892. Translated by Benjamin Jowett. In *The Dialogues of Plato*, 3rd ed., vol. III. Originally composed c. 380 BCE.

Carol Ann Podgorski, Sharon D. Anderson, and Jasneet Parmar. A biopsychosocial-ecological framework for family-framed dementia care. *Frontiers in Psychiatry*, 12: 744806, 2021. doi: 10.3389/fpsyt.2021.744806.

Anneke Poelen, Marieke van Kuppenveld, and Anke Persoon. Nurses' perspectives on shift-to-shift handovers in relation to person-centred nursing home care. *Nursing Open*, 10(8):5035–5043, 2023. doi: 10.1002/nop2.1740.

PointClickCare Technologies Inc. PointClickCare sets industry standard in long-term and post-acute care (LTPAC) with new ONC health IT certification. `https://pointclickcare.com/press-releases/pointclickcare-sets-industry-standard-in-long-term-and-post-acute-care-with-new-onc-health-it-certification/`, February 2024. Press release, February 26, 2024, reporting platform adoption by over 27,000 long-term and post-acute care providers. Accessed 2026-08-03.

PointClickCare Technologies Inc. PointClickCare: Cloud-based healthcare software for long-term and post-acute care. `https://www.pointclickcare.com/`, 2025. Accessed 2025-07-26; URL re-verified live 2026-08-03.

Stefan Popenici. *Artificial Intelligence and Learning Futures: Critical Narratives of Technology and Imagination in Higher Education*. Routledge, 2022. ISBN 978-1-032-20852-7. doi: 10.4324/9781003266563.

Deborah M. Price, Linda Strodtman, Elizabeth Brough, Steven Lonn, and Airong Luo. Digital storytelling: an innovative technological approach to nursing education. *Nurse Educator*, 40(2):66–70, 2015. doi: 10.1097/NNE.0000000000000094.

Isaac Prilleltensky, Geoffrey Nelson, and Leslea Peirson. The role of power and control in children's lives: An ecological analysis of pathways toward wellness, resilience and problems. *Journal of Community & Applied Social Psychology*, 11(2):143–158, 2001. doi: 10.1002/casp.616.

Stephen Prochaska, Julie Vera, Douglas Lew Tan, Ben Yamron, Sylvie Venuto, Amaya Kejriwal, Sarah Chu, and Kate Starbird. Deep storytelling: Collective sensemaking and layers of meaning in U.S. elections. *Proceedings of the ACM on Human-Computer Interaction*, 9(7):1–43, 2025. doi: 10.1145/3757576.

Panagiotis Psomos and Maria Kordaki. Pedagogical analysis of educational digital storytelling environments of the last five years. *Procedia - Social and Behavioral Sciences*, 46:1213–1218, 2012. doi: 10.1016/j.sbspro.2012.05.277.

Sadhana Puntambekar and Roland Hübscher. Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist*, 40(1):1–12, 2005. doi: 10.1207/s15326985ep4001_1.

Sadhana Puntambekar and Janet L. Kolodner. Toward implementing distributed scaffolding: Helping students learn science from design. *Journal of Research in Science Teaching*, 42(2):185–217, 2005. doi: 10.1002/tea.20048.

Gloria Puurveen, Jennifer Baumbusch, and Preet Gandhi. From family involvement to family inclusion in nursing home settings: A critical interpretive synthesis. *Journal of Family Nursing*, 24(1):60–85, 2018. doi: 10.1177/1074840718754314.

Peter Pype, Fien Mertens, Fleur Helewaut, and Demi Krystallidou. Healthcare teams as complex adaptive systems: understanding team behaviour through team members' perception of interpersonal interaction. *BMC Health Services Research*, 18(1):570, 2018. doi: 10.1186/s12913-018-3392-3.

Igor Pyrko, Viktor Dörfler, and Colin Eden. Thinking together: What makes communities of practice work? *Human Relations*, 70(4):389–409, 2017. doi: 10.1177/0018726716661040.

Chia Yi Quah and Kher Hui Ng. A systematic literature review on digital storytelling authoring tool in education: January 2010 to January 2020. *International Journal of Human–Computer Interaction*, 38(9):851–867, 2022. doi: 10.1080/10447318.2021.1972608.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 28492–28518. PMLR, 2023.

Randstad. Workmonitor 2023: Flexible, but stable. Technical report, Randstad N.V., 2023. URL `https://workforceinsights.randstad.com/hubfs/Workmonitor/2023/Randstad_Workmonitor_2023.pdf`. Accessed 2026-08-03.

Cheryl Rathert, Julie Brandt, and Eric S. Williams. Putting the 'patient' in patient safety: a qualitative study of consumer experiences. *Health Expectations*, 15(3):327–336, 2012. doi: 10.1111/j.1369-7625.2011.00685.x.

Rob Reich, Mehran Sahami, and Jeremy M. Weinstein. *System Error: Where Big Tech Went Wrong and How We Can Reboot*. Harper, New York, 2021. ISBN 9780063064881.

R. Colin Reid, Neena L. Chappell, and Jessica A. Gish. Measuring family perceived involvement in individualized long-term care. *Dementia*, 6(1):89–104, 2007. doi: 10.1177/1471301207075640.

Brian J. Reiser. Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, 13(3):273–304, 2004. doi: 10.1207/s15327809jls1303_2.

Hanne Mette Ridder, Julie Kolbe Krøier, Jens Anderson-Ingstrup, and Orii McDermott. Person-attuned musical interactions (PAMI) in dementia care. Complex intervention research for constructing a training manual. *Frontiers in Medicine*, 10:1160588, 2023. doi: 10.3389/fmed.2023.1160588.

Hanne Mette O. Ridder, Brynjulf Stige, Liv Gunnhild Qvale, and Christian Gold. Individual music therapy for agitation in dementia: an exploratory randomized controlled trial. *Aging & Mental Health*, 17(6):667–678, 2013. doi: 10.1080/13607863.2013.790926.

Horst W. J. Rittel and Melvin M. Webber. Dilemmas in a general theory of planning. *Policy Sciences*, 4(2):155–169, 1973. doi: 10.1007/BF01405730.

Joanne Roberts. Limits to communities of practice. *Journal of Management Studies*, 43 (3):623–639, 2006. doi: 10.1111/j.1467-6486.2006.00618.x.

Judy Robertson and Maurits Kaptein. An introduction to modern statistical methods in HCI. In Judy Robertson and Maurits Kaptein, editors, *Modern Statistical Methods for HCI*, Human–Computer Interaction Series, pages 1–14. Springer, Cham, Switzerland, 2016. ISBN 978-3-319-26631-2. doi: 10.1007/978-3-319-26633-6_1.

Bernard Robin. The educational uses of digital storytelling. In C. Crawford, R. Carlsen, K. McFerrin, J. Price, R. Weber, and D. Willis, editors, *Proceedings of SITE 2006–Society for Information Technology & Teacher Education International Conference*, pages 709–716, Orlando, FL, USA, 2006. Association for the Advancement of Computing in Education (AACE). URL `https://www.learntechlib.org/p/22129/`.

Carl R. Rogers. *On Becoming a Person: A Therapist's View of Psychotherapy*. Houghton Mifflin, Boston, 1961.

Eka Roivainen. Gender differences in processing speed: A review of recent research. *Learning and Individual Differences*, 21(2):145–149, 2011. doi: 10.1016/j.lindif.2010.11.021.

Joshua M. Rosenberg and Matthew J. Koehler. Context and technological pedagogical content knowledge (TPACK): A systematic review. *Journal of Research on Technology in Education*, 47(3):186–210, 2015. doi: 10.1080/15391523.2015.1052663.

Sherry Ruan, Jiayu He, Rui Ying, Jonathan Burkle, Dunia Hakim, Anna Wang, Yufeng Yin, Lily Zhou, Qianyao Xu, Abdallah AbuHashem, Griffin Dietz, Elizabeth L. Murnane, Emma Brunskill, and James A. Landay. Supporting children's math learning with feedback-augmented narrative technology. In *Proceedings of the Interaction Design and Children Conference*, IDC '20, pages 567–580, New York, NY, USA, 2020. Association for Computing Machinery. doi: 10.1145/3392063.3394400.

Daniel M. Russell, Mark J. Stefik, Peter Pirolli, and Stuart K. Card. The cost structure of sensemaking. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, CHI '93, pages 269–276, New York, NY, USA, 1993. Association for Computing Machinery. doi: 10.1145/169059.169209.

Daniel M. Russell, Laura Koesten, Aniket Kittur, and Regina Schuster. Sensemaking and AI 2026: Uses, behaviors, design, and recommendations. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems*, pages 1–4, 2026. doi: 10.1145/3772363.3779006.

Alaa Sadik. Digital storytelling: A meaningful technology-integrated approach for engaged student learning. *Educational Technology Research and Development*, 56(4):487–506, 2008. doi: 10.1007/s11423-008-9091-8.

Francesco Salvi, Manoel Horta Ribeiro, Riccardo Gallotti, and Robert West. On the conversational persuasiveness of GPT-4. *Nature Human Behaviour*, 9(8):1645–1653, 2025. doi: 10.1038/s41562-025-02194-6.

Ravi S. Sandhu, Edward J. Coyne, Hal L. Feinstein, and Charles E. Youman. Role-based access control models. *Computer*, 29(2):38–47, 1996. doi: 10.1109/2.485845.

Hatice Çıralı Sarıca and Yasemin Koçak Usluel. The effect of digital storytelling on visual memory and writing skills. *Computers & Education*, 94:298–309, 2016. doi: 10.1016/j.compedu.2015.11.016.

John R. Savery. Overview of problem-based learning: Definitions and distinctions. In Andrew Walker, Heather Leary, Cindy E. Hmelo-Silver, and Peggy A. Ertmer, editors, *Essential Readings in Problem-Based Learning: Exploring and Extending the Legacy of Howard S. Barrows*, pages 5–16. Purdue University Press, West Lafayette, IN, 2015. doi: 10.2307/j.ctt6wq6fh.6.

Tanja Schultz, Felix Putze, Timo Schulze, Lars Steinert, Ralf Mikut, Wolfgang Doneit, Andreas Kruse, Anamaria Depner, Ingo Franz, Marc Aurel Engels, Philipp Gaerte, Sebastian Jünger, Rene Linden, Christof Ziegler, Michael Ricken, Todor Dimitrov, Joachim Herzig, Irene Maucher, Keni Bernardin, Tobias Gehrig, Jana Lohse, Kristina Glesing, Monika Fischer, and Clarissa Simon. I-CARE: Ein Mensch-Technik Interaktionssystem zur Individuellen Aktivierung von Menschen mit Demenz. In Susanne Boll, Andreas Hein, Wilko Heuten, and Karin Wolf-Ostermann, editors, *Zukunft der Pflege: Tagungsband der 1. Clusterkonferenz 2018 – Innovative Technologien für die Pflege*, pages 94–99, Oldenburg, Germany, 2018. BIS-Verlag. ISBN 978-3-8142-2367-4.

Andrew D. Selbst, danah boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*, pages 59–68, New York, NY, USA, 2019. ACM. doi: 10.1145/3287560.3287598.

Amartya Sen. *Development as Freedom*. Oxford University Press, Oxford, 1999. ISBN 9780198297581.

Sefik Ilkin Serengil and Alper Ozpinar. HyperExtended LightFace: A facial attribute analysis framework. In *2021 International Conference on Engineering and Emerging Technologies (ICEET)*, pages 1–4, Istanbul, Turkey, 2021. IEEE. doi: 10.1109/ICEET53442.2021.9659697.

Nalin K. Sharda. Authoring educational multimedia content using learning styles and story telling principles. In *Proceedings of the International Workshop on Educational Multimedia and Multimedia Education (Emme '07)*, pages 93–102, New York, NY, USA, 2007. ACM. doi: 10.1145/1290144.1290160.

Ashish Sharma, Inna W. Lin, Adam S. Miner, David C. Atkins, and Tim Althoff. Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nature Machine Intelligence*, 5(1):46–57, 2023. doi: 10.1038/s42256-022-00593-2.

Nikita Sharma, Jeroen Klein Brinke, J. E. W. C. van Gemert-Pijnen, and L. M. A. Braakman-Jansen. Implementation of unobtrusive sensing systems for older adult care: Scoping review. *JMIR Aging*, 4(4):e27862, 2021. doi: 10.2196/27862.

Susan W. Sherman. Multiple choice test bias uncovered by use of an "I Don't Know" alternative, 1976. URL `https://eric.ed.gov/?id=ED121824`. ERIC Document ED121824. Accessed: 2026-08-03.

Meilin Shi, Guiling Wang, Shaohua Zhang, Jinlei Jiang, and Yong Xiang. AISCW: Artificial intelligence supported cooperative work the new frontier of CSCW. In *Computer Supported Cooperative Work and Social Computing (ChineseCSCW 2024)*, volume 2343 of *Communications in Computer and Information Science*, pages 3–17. Springer Nature Singapore, 2025. doi: 10.1007/978-981-96-2373-0_1.

Regina A. Shih, Thomas W. Concannon, Jodi L. Liu, and Esther M. Friedman. Improving dementia long-term care: A policy blueprint. *RAND Health Quarterly*, 4(2):2, 2014. doi: 10.7249/RHQ-04-02-02.

Herbert A. Simon. A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1):99–118, 1955. doi: 10.2307/1884852.

Anjali Singh, Christopher Brooks, and Shayan Doroudi. Learnersourcing in theory and practice: Synthesizing the literature and charting the future. In *Proceedings of the Ninth ACM Conference on Learning @ Scale (L@S '22)*, pages 234–245, New York, NY, USA, 2022. Association for Computing Machinery. doi: 10.1145/3491140.3528277.

Anjali Singh, Christopher Brooks, Xu Wang, Warren Li, Juho Kim, and Deepti Wilson. Bridging learnersourcing and AI: Exploring the dynamics of student-AI collaborative feedback generation. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*, LAK '24, pages 742–748, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400716188. doi: 10.1145/3636555.3636853. Previously arXiv:2311.12148 (2023); the preprint lists the last author as Deepti Pandey.

Najat Smeda, Eva Dakich, and Nalin Sharda. Developing a framework for advancing e-learning through digital storytelling. In Miguel Baptista Nunes and Maggie McPherson, editors, *Proceedings of the IADIS International Conference on e-Learning*, pages 169–176, Freiburg, Germany, 2010. International Association for Development of the Information Society (IADIS).

Mark Snowden, Kersten Sato, and Peter Roy-Byrne. Assessment and treatment of nursing home residents with depression or behavioral symptoms associated with dementia: a review of the literature. *Journal of the American Geriatrics Society*, 51(9):1305–1317, 2003. doi: 10.1046/j.1532-5415.2003.51417.x.

Carolyn Snyder. *Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces*. Morgan Kaufmann, San Francisco, CA, 2003. ISBN 9781558608702.

Lídia Sousa, Maria J. Neves, Bárbara Moura, Justine Schneider, and Lia Fernandes. Music-based interventions for people living with dementia, targeting behavioral and psycholog-

ical symptoms: A scoping review. *International Journal of Geriatric Psychiatry*, 36(11): 1664–1690, 2021. doi: 10.1002/gps.5568.

Susan Leigh Star and Karen Ruhleder. Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, 7(1):111–134, 1996. doi: 10.1287/isre.7.1.111.

Amy J. Starmer, Nancy D. Spector, Rajendu Srivastava, Daniel C. West, Glenn Rosenbluth, April D. Allen, Elizabeth L. Noble, Lisa L. Tse, Anuj K. Dalal, Carol A. Keohane, Stuart R. Lipsitz, Jeffrey M. Rothschild, Matthew F. Wien, Catherine S. Yoon, Katherine R. Zigmont, Karen M. Wilson, Jennifer K. O'Toole, Lauren G. Solan, Megan Aylor, Zia Bismilla, Maitreya Coffey, Sanjay Mahant, Rebecca L. Blankenburg, Lauren A. Destino, Jennifer L. Everhart, Shilpa J. Patel, James F. Bale, Jr., Jaime B. Spackman, Adam T. Stevenson, Sharon Calaman, F. Sessions Cole, Dorene F. Balmer, Jennifer H. Hepps, Joseph O. Lopreiato, Clifton E. Yu, Theodore C. Sectish, Christopher P. Landrigan, and the I-PASS Study Group. Changes in medical errors after implementation of a handoff program. *New England Journal of Medicine*, 371(19):1803–1812, 2014. doi: 10.1056/NEJMsa1405556.

Lars Steinert, Felix Putze, Dennis Küster, and Tanja Schultz. Towards engagement recognition of people with dementia in care settings. In *Proceedings of the 2020 International Conference on Multimodal Interaction*, ICMI '20, page 558–565, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450375818. doi: 10.1145/3382507.3418856. URL https://doi.org/10.1145/3382507.3418856.

Ileana Stigliani and Davide Ravasi. Organizing thoughts and connecting brains: Material practices and the transition from individual to group-level prospective sensemaking. *Academy of Management Journal*, 55(5):1232–1259, 2012. doi: 10.5465/amj.2010.0890.

Jiahong Su, Davy Tsz Kit Ng, and Samuel Kai Wah Chu. Artificial intelligence (AI) literacy in early childhood education: The challenges and opportunities. *Computers and Education: Artificial Intelligence*, 4:100124, 2023. doi: 10.1016/j.caeai.2023.100124.

Ponnusamy Subramaniam and Bob Woods. The impact of individual reminiscence therapy for people with dementia: systematic review. *Expert Review of Neurotherapeutics*, 12 (5):545–555, 2012. doi: 10.1586/ern.12.35.

Sangho Suh, Martinet Lee, Gracie Xia, and Edith Law. Coding Strip: A pedagogical tool for teaching and learning programming concepts through comics. In *2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 1–10, Dunedin, New Zealand, 2020. IEEE. doi: 10.1109/VL/HCC50065.2020.9127262.

Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. Sensecape: Enabling multilevel exploration and sensemaking with large language models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, 2023. doi: 10.1145/3586183.3606756.

Nicole Sultanum, Michael Brudno, Daniel Wigdor, and Fanny Chevalier. More text please! Understanding and supporting the use of visualization for clinical text overview. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 1–13, Montreal QC, Canada, 2018. Association for Computing Machinery. doi: 10.1145/3173574.3173996.

George Sun and Yi-Hui Zhou. AI in healthcare: navigating opportunities and challenges in digital communication. *Frontiers in Digital Health*, 5:1291132, 2023. doi: 10.3389/fdgth.2023.1291132.

Ruth Sylvester and Wendy-lou Greenidge. Digital storytelling: Extending the potential for struggling writers. *The Reading Teacher*, 63(4):284–295, 2009. doi: 10.1598/RT.63.4.3.

Joanna Szurmak and Mindy Thuna. Tell me a story: The use of narrative as a tool for instruction. In *Imagine, Innovate, Inspire: The Proceedings of the ACRL 2013 Conference*, pages 546–552, Indianapolis, IN, 2013. Association of College and Research Libraries. URL https://www.ala.org/sites/default/files/acrl/content/conferences/confsandpreconfs/2013/papers/SzurmakThuna_TellMe.pdf.

Iris Tabak. Synergy: A complement to emerging patterns of distributed scaffolding. *Journal of the Learning Sciences*, 13(3):305–335, 2004. doi: 10.1207/s15327809jls1303_3.

Jane Tan, Shiou-Liang Wee, Pei Shi Yeo, Juliet Choo, Michele Ritholz, and Philip Yap. A new music therapy engagement scale for persons with dementia. *International Psychogeriatrics*, 31(1):49–58, 2019. doi: 10.1017/S1041610218000509.

Karin Tengler, Oliver Kastner-Hauler, and Barbara Sabitzer. Enhancing computational thinking skills using robots and digital storytelling. In *Proceedings of the 13th International Conference on Computer Supported Education (CSEDU)*, volume 1, pages 157–164. SciTePress, 2021. doi: 10.5220/0010477001570164.

Danielle R. Thomas, Shivang Gupta, and Kenneth R. Koedinger. Comparative analysis of learnersourced human-graded and AI-generated responses for autograding online tutor lessons. In *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky (AIED 2023)*, volume 1831 of *Communications in Computer and Information Science*, pages 714–719, Tokyo, Japan, 2023. Springer. doi: 10.1007/978-3-031-36336-8_110.

Melika Torabgar, Mathieu Figeys, Shaniff Esmail, Eleni Stroulia, and Adriana M Ríos Rincón. Machine learning analysis of engagement behaviors in older adults with dementia playing mobile games: Exploratory study. *JMIR Serious Games*, 13:e54797, 2025. doi: 10.2196/54797.

David Touretzky, Christina Gardner-McCune, Fred Martin, and Deborah Seehorn. Envisioning AI for K-12: What should every child know about AI? *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):9795–9799, 2019. doi: 10.1609/aaai.v33i01.33019795.

Bernie Trilling and Charles Fadel. *21st Century Skills: Learning for Life in Our Times*. Jossey-Bass, San Francisco, CA, 2009. ISBN 9780470475386.

Joan C Tronto. An ethic of care. *Generations: Journal of the American Society on Aging*, 22(3):15–20, 1998. PMID: 12785337.

Joan C. Tronto. *Moral Boundaries: A Political Argument for an Ethic of Care*. Routledge, 2020. ISBN 9781003070672. doi: 10.4324/9781003070672. eBook edition; originally published 1993 by Routledge, New York.

Lorainne Tudor Car, Dhakshenya Ardhithy Dhinagaran, Bhone Myint Kyaw, Tobias Kowatsch, Shafiq Joty, Yin-Leng Theng, and Rifat Atun. Conversational agents in health care: Scoping review and conceptual analysis. *Journal of Medical Internet Research*, 22 (8):e17158, 2020. doi: 10.2196/17158.

Brad R. Turner. Narrative affordances: What stories can and cannot do. *Academy of Management Annals*, 19(2):522–564, 2025. doi: 10.5465/annals.2023.0101.

John R. Turner, Jeff Allen, Suliman Hawamdeh, and Gujjula Mastanamma. The multifaceted sensemaking theory: A systematic literature review and content analysis on sensemaking. *Systems*, 11(3):145, 2023. doi: 10.3390/systems11030145.

UNESCO Institute for Lifelong Learning. *Making Lifelong Learning a Reality: A Handbook*. UNESCO Institute for Lifelong Learning, Hamburg, Germany, 2022. ISBN 978-92-820-1238-3.

Queen Utley-Smith, Cathleen S Colón-Emeric, Deborah Lekan-Rutledge, Natalie Ammarell, Donald Bailey, Kirsten Corazzini, Mary L Piven, and Ruth A Anderson. Staff perceptions of staff-family interactions in nursing homes. *Journal of Aging Studies*, 23 (3):168–177, 2009. doi: 10.1016/j.jaging.2007.11.003.

Janneke van de Pol, Monique Volman, and Jos Beishuizen. Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, 22(3):271–296, 2010. doi: 10.1007/s10648-010-9127-6.

Jenny T van der Steen, Johannes C van der Wouden, Abigail M Methley, Hanneke J A Smaling, Annemieke C Vink, and Manon S Bruinsma. Music-based therapeutic interventions for people with dementia. *Cochrane Database of Systematic Reviews*, 2025(3): CD003477, 2025. doi: 10.1002/14651858.CD003477.pub5.

Ester van Laar, Alexander J. A. M. van Deursen, Jan A. G. M. van Dijk, and Jos de Haan. The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior*, 72:577–588, 2017. doi: 10.1016/j.chb.2017.03.010.

Alexander Vedernikov, Zhaodong Sun, Virpi-Liisa Kykyri, Mikko Pohjola, Miriam Nokia, and Xiaobai Li. Analyzing participants' engagement during online meetings using unsupervised remote photoplethysmography with behavioral features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 389–399, Seattle, WA, USA, 2024. IEEE. doi: 10.1109/CVPRW63382.2024.00044.

Viswanath Venkatesh, James Y. L. Thong, and Xin Xu. Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 36(1):157–178, 2012. doi: 10.2307/41410412.

Myrra Vernooij-Dassen, Emmelyne Vasse, Sytse Zuidema, Jiska Cohen-Mansfield, and Wendy Moyle. Psychosocial interventions for dementia patients in long-term care. *International Psychogeriatrics*, 22(7):1121–1128, 2010. doi: 10.1017/S1041610210001365.

Dianna Vidas, Zara Thompson, Ryan M. Kelly, Jenny Waycott, Jeanette Tamplin, Tanara Vieira Sousa, Lars Kulik, Amit Lampit, Nicola T. Lautenschlager, and Felicity A. Baker. Family caregiver experiences of using a mobile app for music-based training to support dementia care. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, pages 1–16, New York, NY, USA, 2024. Association for Computing Machinery. doi: 10.1145/3613904.3642308.

Feliciano Villar, Rodrigo Serrat, and Stephany Bravo-Segal. Giving them a voice: Challenges to narrative agency in people with dementia. *Geriatrics*, 4(1):20, 2019. doi: 10.3390/geriatrics4010020.

Marianna Vivitsou, Veera Kallunki, Hannele Niemi, Johanna Penttilä, and Vilhelmiina Harju. Student-driven knowledge creation through digital storytelling. In Hannele Niemi and Jiyou Jia, editors, *New Ways to Teach and Learn in China and Finland: Crossing Boundaries with Technology*, pages 35–53. Peter Lang, Frankfurt am Main, 2016. doi: 10.3726/978-3-631-69873-0.

Ludwig von Bertalanffy. *General System Theory: Foundations, Development, Applications*. George Braziller, New York, 1968.

Jakob Voß. Measuring Wikipedia. In Peter Ingwersen and Birger Larsen, editors, *Proceedings of ISSI 2005: The 10th International Conference of the International Society for Scientometrics and Informetrics*, volume 1, pages 221–231, Stockholm, Sweden, 2005. Karolinska University Press.

Lev S. Vygotsky. *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, Cambridge, MA, 1978. Edited by Cole, Michael, John-Steiner, Vera, Scribner, Sylvia, Souberman, Ellen.

Alex Waddell, Joshua Paolo Seguin, Ling Wu, Peta Stragalinos, Joe Wherton, Jessica L. Watterson, Christopher Owen Prawira, Patrick Olivier, Victoria Manning, Dan Lubman, and Jasmin Grigg. Leveraging implementation science in human-centred design for digital health. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, pages 1–17, New York, NY, USA, 2024. Association for Computing Machinery. doi: 10.1145/3613904.3642161.

Gregory M. Walton and Geoffrey L. Cohen. A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology*, 92(1):82–96, 2007. doi: 10.1037/0022-3514.92.1.82.

Lu Wang, Diva Smriti, Hao Yuan, and Jina Huh-Yoo. Artificial intelligence systems for supporting informal caregivers of people living with Alzheimer's disease or related de-

mentias: A systematic review. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–11, New York, NY, USA, 2024a. Association for Computing Machinery. doi: 10.1145/3613905.3650846.

Tianchong Wang and Eric C. K. Cheng. Towards a tripartite research agenda: A scoping review of artificial intelligence in education research. In Eric C. K. Cheng, Rekha B. Koul, Tianchong Wang, and Xinguo Yu, editors, *Artificial Intelligence in Education: Emerging Technologies, Models and Applications (Proceedings of 2021 2nd International Conference on Artificial Intelligence in Education Technology)*, volume 104 of *Lecture Notes on Data Engineering and Communications Technologies*, pages 3–24, Singapore, 2022. Springer. doi: 10.1007/978-981-16-7527-0_1.

Yihan Wang, Shu Liu, Alanna G. Spiteri, Andrew Liem Hieu Huynh, Chenyin Chu, Colin L. Masters, Benjamin Goudey, Yijun Pan, and Liang Jin. Understanding machine learning applications in dementia research and clinical practice: a review for biomedical scientists and clinicians. *Alzheimer's Research & Therapy*, 16(1):175, 2024b. doi: 10.1186/s13195-024-01540-6.

Kerri Wazny. "Crowdsourcing" ten years in: A review. *Journal of Global Health*, 7(2): 020601, 2017. doi: 10.7189/jogh.07.020601.

Karl E. Weick. *The Social Psychology of Organizing*. Addison-Wesley, Reading, MA, 1969.

Karl E. Weick. *The Social Psychology of Organizing*. Addison-Wesley, Reading, MA, 2nd edition, 1979.

Karl E. Weick. *Sensemaking in Organizations*. Sage Publications, Thousand Oaks, CA, 1995.

Karl E. Weick, Kathleen M. Sutcliffe, and David Obstfeld. Organizing and the process of sensemaking. *Organization Science*, 16(4):409–421, 2005. doi: 10.1287/orsc.1050.0133.

Bryan J. Weiner, Cara C. Lewis, Cameo Stanick, Byron J. Powell, Caitlin N. Dorsey, Alecia S. Clary, Marcella H. Boynton, and Heather Halko. Psychometric assessment of three newly developed implementation outcome measures. *Implementation Science*, 12 (1):108, 2017. doi: 10.1186/s13012-017-0635-3.

Sarah Weir, Juho Kim, Krzysztof Z. Gajos, and Robert C. Miller. Learnersourcing subgoal labels for how-to videos. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, pages 405–416, New York, NY, USA, 2015. Association for Computing Machinery. doi: 10.1145/2675133.2675219.

Mark Weiser and John Seely Brown. The coming age of calm technology. In Peter J. Denning and Robert M. Metcalfe, editors, *Beyond Calculation: The Next Fifty Years of Computing*, pages 75–85. Springer, New York, NY, 1997. doi: 10.1007/978-1-4612-0685-9_6.

Etienne Wenger. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press, Cambridge, UK, 1998. doi: 10.1017/CBO9780511803932.

Etienne Wenger, Richard McDermott, and William M. Snyder. *Cultivating Communities of Practice: A Guide to Managing Knowledge*. Harvard Business School Press, Boston, MA, 2002.

Etienne Wenger-Trayner, Mark Fenton-O'Creevy, Steven Hutchinson, Chris Kubiak, and Beverly Wenger-Trayner, editors. *Learning in Landscapes of Practice: Boundaries, Identity, and Knowledgeability in Practice-Based Learning*. Routledge, Abingdon, UK, 2015. doi: 10.4324/9781315777122.

Eunike Wetzel, Jan R. Böhnke, and Anna Brown. Response biases. In Frederick T. L. Leong, Dave Bartram, Fanny Cheung, Kurt F. Geisinger, and Dragos Iliescu, editors, *The ITC International Handbook of Testing and Assessment*, pages 349–363. Oxford University Press, New York, 2016. ISBN 9780199356942. doi: 10.1093/med:psych/9780199356942.003.0024.

Michael White. *Maps of Narrative Practice*. W. W. Norton, New York, NY, 2007.

Jeffrey L. Whitten, Lonnie D. Bentley, and Thomas I. M. Ho. *Systems Analysis & Design Methods*. Times Mirror/Mosby College Publishing, St. Louis, MO, 1986. ISBN 9780801654640.

Jaime Williams, Thomas Hadjistavropoulos, Omeed O. Ghandehari, Xue Yao, and Lisa Lix. An evaluation of a person-centred care programme for long-term care facilities. *Ageing & Society*, 35(3):457–488, 2015. doi: 10.1017/S0144686X13000743.

Kristine N. Williams, Yelena Perkhounkova, Ying-Ling Jao, Ann Bossen, Maria Hein, Sophia Chung, Anne Starykowicz, and Margaret Turk. Person-centered communication for nursing home residents with dementia: Four communication analysis methods. *Western Journal of Nursing Research*, 40(7):1012–1031, 2018. doi: 10.1177/0193945917697226.

Vicki Williams. Online learning readiness questionnaire. The Pennsylvania State University, 2001. URL `https://pennstate.qualtrics.com/jfe/form/SV_7QCNUPsyH9f012B`. Original page at `https://tutorials.istudy.psu.edu/learningonline/ORQ/ORQ.htm` no longer live; archived at `https://web.archive.org/web/20220901071414/https://tutorials.istudy.psu.edu/learningonline/ORQ/ORQ.htm` (snapshot 2022-09-01). Year as attributed in prior literature. Accessed: 2026-08-03.

David Wood, Jerome S. Bruner, and Gail Ross. The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2):89–100, 1976. doi: 10.1111/j.1469-7610.1976.tb00381.x.

World Health Organization. WHO global strategy on people-centred and integrated health services: interim report. Technical Report WHO/HIS/SDS/2015.6, World Health Organization, Geneva, 2015. URL `https://iris.who.int/handle/10665/155002`.

World Health Organization. *Global status report on the public health response to dementia*. World Health Organization, Geneva, 2021. ISBN 978-92-4-003324-5.

World Health Organization. Dementia. Fact sheet, 2023. URL https://www.who.int/news-room/fact-sheets/detail/dementia. Fact sheet dated 15 March 2023. Live page since updated (3 July 2026 edition as of access); 2023 version archived at https://web.archive.org/web/20230922161917/https://www.who.int/news-room/fact-sheets/detail/dementia. Accessed: 2026-08-03.

Scott D. Wurdinger and Julie A. Carlson. *Teaching for Experiential Learning: Five Approaches That Work*. Rowman & Littlefield Education, Lanham, MD, 2010. ISBN 9781607093688.

Qian Yang, Aaron Steinfeld, and John Zimmerman. Unremarkable AI: Fitting intelligent decision support into critical, clinical decision-making processes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–11. ACM, 2019. doi: 10.1145/3290605.3300468.

Ya-Ting C. Yang and Wan-Chi I. Wu. Digital storytelling for enhancing student academic achievement, critical thinking, and learning motivation: A year-long experimental study. *Computers & Education*, 59(2):339–352, 2012. doi: 10.1016/j.compedu.2011.12.012.

Nuray Parlak Yılmaz. Learning experiences of prospective teachers in a digital storytelling tool called "Toondoo": a case study. In Recep Efe, Irina Koleva, Emin Atasoy, and Violeta Kotseva, editors, *Current Trends in Educational Sciences*, pages 447–454. St. Kliment Ohridski University Press, Sofia, 2017.

Robert K. Yin. *Case Study Research and Applications: Design and Methods*. SAGE Publications, Thousand Oaks, CA, 6th edition, 2018.

Pelin Yuksel-Arslan, Soner Yildirim, and Bernard Ross Robin. A phenomenological study: teachers' experiences of using digital storytelling in early childhood education. *Educational Studies*, 42(5):427–445, 2016. doi: 10.1080/03055698.2016.1195717.

Shumenghui Zhai, Frances Chu, Minghui Tan, Nai-Ching Chi, Teresa Ward, and Weichao Yuwen. Digital health interventions to support family caregivers: An updated systematic review. *Digital Health*, 9:20552076231171967, 2023. doi: 10.1177/20552076231171967.

Tao Zhou and Songtao Li. Understanding user switch of information seeking: From search engines to generative AI. *Journal of Librarianship and Information Science*, 58(1):696–708, 2026. doi: 10.1177/09610006241244800.

Xiaofei Zhou, Jessica Van Brummelen, and Phoebe Lin. Designing AI learning experiences for K-12: Emerging works, future opportunities and a design framework, 2020. URL https://arxiv.org/abs/2009.10228. arXiv preprint arXiv:2009.10228.