Making Visible, Making Invisible: How an AI Scribe Reshapes Documentation Authority in Social Work

Eunki Joung [1] **Hong Chul Nam** [2] **Geonhwi Hwang** [3] **Moohyun Lee** [1] **Hyun Kwon** [1]

Abstract

AI scribes have shown promise in reducing clinical documentation burden and are now extending into social work. Social work documentation, however, serves a distinct function: it mediates between individual clients and the welfare regime, determining eligibility and shaping whose needs are recognized. We designed and deployed *Care Fish*, an AI scribe technology probe for in-home older adult care in South Korea, with four social workers. Participants reported that Care Fish improved documentation efficiency by surfacing information missed during client interviews, and they identified and corrected its errors during a re-entry step into their existing systems. The deployment also surfaced shifts in documentation practice. The recording and AI-generated citations acquired a form of authority that workers compared against their own observations, reinforcing an existing epistemological hierarchy in which institutional ways of knowing outrank both worker observation and client speech. The expectation of being recorded changed how both clients and workers spoke during case interviews. We discuss how these shifts bear on workers' professional discretion and clients' self-presentation and consider design directions that make the AI scribe's own perspective visible to the workers who use it.

1. Introduction

Documentation is an essential task of social work, where social workers recognize clients' needs and match them with appropriate public services. It serves as a translation between the differing information needs of public institutions, service clients, managers, and social workers (Sum et al., 2023). For instance, a client's difficulty in preparing meals might be translated into eligibility for a meal-support service. Documentation is also critical for ensuring quality and legality of core social work tasks such as assessment, planning, and service delivery (Reamer, 2005).

Social workers struggle with a heavy burden of documentation, which is a main source of worker stress and burnout (Nelson et al., 2024; Gondimalla et al., 2024). Workers often face a scarcity of resources while serving many clients (Lee and Kwon, 2020), and the documentation burden can feel detached from their core responsibilities (Yoon and Yoo, 2023; Jo et al., 2025), even though the produced documents are often still insufficient (Kuorikoski, 2024).

AI scribes have recently gained popularity in clinical contexts to address similar challenges. These systems combine automatic speech recognition, natural language processing, and a review interface to support documentation from recorded conversations. Studies report that AI scribes reduce documentation burden and improve clinician-patient interaction (Cao et al., 2024; Tierney et al., 2024; Nair et al., 2026). Prior work on clinical AI scribes has systematically evaluated their accuracy, consistency, and effects on provider–patient interactions (Sasseville et al., 2025; Lee et al., 2024). Through prototyping, researchers have also examined how physicians perceive current and future AI scribes and how these tools reshape clinical workflows (Li et al., 2021). However, the distinctive context of social services remains largely unexamined.

Social work documentation shares surface similarities with clinical documentation: both involve translating professional conversations into institution-required formats. Social work case documents, however, serve as the interface between individual clients and the welfare regime, determining eligibility, allocating resources, and shaping whose needs are recognized. How an AI scribe translates conversations into these documents carries implications beyond accuracy or efficiency.

Research on AI scribes in non-clinical social work settings remains limited. Reported deployments concentrate in the UK public sector: most prominently Beam's Magic Notes,

[1]Korea Military Academy, Seoul, Republic of Korea [2]ETH Zurich, Zurich, Switzerland [3]Hokkaido University, Sapporo, Japan. Correspondence to: Hyun Kwon <hkwon.cs@gmail.com>.

Trustworthy AI for Good Workshop at the 43rd International Conference on Machine Learning, Seoul, South Korea, 2026. Copyright 2026 by the author(s).

a domain-specific tool now in use across some local authorities, alongside more ad-hoc use of general-purpose systems such as Microsoft Copilot (Meers et al., 2026). While adoption has scaled rapidly, empirical evaluation is still at an early stage, and reported public attitudes are a mix of enthusiasm and resistance. Alongside this early-stage adoption, a broader literature on AI in public institutions warns that automated systems can reshape the dynamics of street-level bureaucracy (Alkhatib and Bernstein, 2019) and reinforce existing inequalities (Eubanks, 2018). How these dynamics play out in AI scribes for social work, however, remains empirically open.

Given this gap, we designed and deployed *Care Fish*, an AI scribe technology probe for in-home older adult care in South Korea. We use the term *client* to refer to the older adult receiving care services, and *social worker* (or *worker*) to refer to the professional who conducts home visits and produces case documents. A *case interview* is a structured conversation between worker and client aimed at assessing needs and matching the client with appropriate services.

We ask three research questions:

1. How do workers perceive an AI scribe for social work documentation?
2. How do workers' documentation practices change when such a system is introduced?
3. What design considerations follow for AI scribes in this setting?

Through a two-week field deployment with three participants and one supplementary lab session, followed by semi-structured interviews, we make three contributions: (1) Care Fish, a technology probe designed through formative fieldwork; (2) empirical findings on how social workers use and negotiate with an AI scribe; and (3) a discussion of how AI scribes can rearrange the distribution of epistemic authority in social work documentation.

2. Design of Technology Probe

2.1. Informing the Design

We visited two in-home older adult care centers, where we had conversations with managers and care workers and observed their documentation practices. Combined with a literature review, these visits identified four design opportunities.

Excessive client-worker ratio. On average, 2–3 social workers care for approximately 50 clients per worker, among which about 10 are more intensively tracked. A manager noted that case work is often postponed because it is less visible than other tasks such as benefit payments and meal delivery, and understanding clients' needs is done at a bare minimum until annual audits make it visible.

Extensive paperwork. Approximately twenty case documents per client are filled out after the first case interview, and similar requirements are repeated annually. These documents are essential for case meetings, service planning, audits, and performance management. Translating a case interview that sometimes exceeds two hours into diverse formal documents can take more than half a day.

Fragmented information systems. Workers produce case documents using handwriting, word processors, and an institutional web database, supplemented by governmental databases, spreadsheets, and presentation software. They struggle to operate these systems simultaneously, leading to inefficiency and incompatibilities. This pattern is consistent with what Voida et al. (2011) called "homebrew databases," the messy, locally assembled set of information systems of resource-constrained social service organizations.

Diverse perspectives. Consistent with Gondimalla et al. (2024), we repeatedly observed that social workers bring diverse perspectives to case documentation. They sometimes have to coordinate conflicting perspectives on defining client needs, in which what clients explicitly want differs from what workers assess they need. Case documents accordingly contain many subjective and open-ended fields.

2.2. System Design

Care Fish is a case note assistant that supports documentation by transcribing audio recordings of client-worker conversations and populating institution-required case documents. We designed it around two approaches.

Automated form filling. Care Fish fills in eleven institution-required case documents based on transcriptions. These belong to two types: evaluation forms (measuring clients' daily living abilities and mental health) and case logs (recording histories of assessment activities, service provision, and case meetings). Users can record or upload audio, then generate multiple document types from a single recording, as shown in Fig. 1.

Evidence and explanation. For each entry of an AI-generated draft, Care Fish provides *evidence* (exact quotes from the transcript grounding the suggestion) and *explanation* (a rationale connecting the evidence to the suggestion). Evidence appears as a pop-up citing transcript excerpts highlighted within the full transcript (Nguyen et al., 2024). For evaluation forms, where workers reported greater need for perspective coordination, both explanation and evidence are displayed; for other forms, only evidence is shown.

The system was built with React and Gemini 2.5 Flash, with CLOVA Speech for transcription. For privacy, the backend

ran locally on each worker's machine; client data was sent only to paid API endpoints that do not use data for training, and was invisible to the researchers.

3. Method

We deployed Care Fish as a technology probe (Hutchinson et al., 2003) to explore what it means to introduce AI scribes into social work documentation. Technology probes prioritize situated understanding over statistical generalization, making them suited for studying how people adopt and make sense of new technologies in their work.

3.1. Study Design

The study proceeded in three phases: (1) a 40-minute introductory session explaining the system, study procedure, and warnings for safer use; (2) a probe usage period: a 14-day field study (3 participants) or a 20-minute lab session via Zoom (1 participant who could not participate in the field deployment due to technical issues); and (3) a 1-hour follow-up semi-structured interview.

We presented participants with warning messages about potential problems, including a general warning ("Care Fish AI is not perfect. ALWAYS REVIEW BY YOUR OWN.") and specific examples of problematic predictions: hallucination that mixes facts and fabrications, and bias reflecting societal conditions embedded in the model.

3.2. Participants

We recruited four social workers (2 female, 2 male; mean age 35.3) from in-home older adult care centers. Mean social work experience was 5.8 years (range: 4–9); mean experience in in-home older adult care was 2.2 years (range: 4 months–5 years). All are current practitioners conducting case management for older adults living at home.

3.3. Data Collection and Analysis

We conducted semi-structured interviews individually, except for two workers who participated together due to shared experience. Interviews lasted approximately 60 minutes and covered how workers used Care Fish, perceptions of each feature, worker agency, navigation of diverse case work approaches, and trust and privacy concerns. At Day 7, we sent check-up messages to assess engagement.

We analyzed interview transcripts through thematic analysis. We iteratively coded the data, identifying descriptive themes close to participants' accounts of their practice, and organized them into three groups: how Care Fish was integrated into workers' documentation work, where it created friction with existing practices, and how it changed what was recorded and how it was valued. The study received IRB approval, and informed consent was obtained from all participants. Recordings and documents were saved locally on workers' machines and not submitted to the researchers.

Table 1. Participants' engagement with Care Fish.

	Type	Rec.	Docs	Forms
P1	Field	4	8	F2, F8
P2	Field	4	44	F1–F11
P3	Field	1	10	—
P4	Lab	—	—	—

All field study participants generated at least 8 documents, exceeding the requested minimum of 3. P2 produced 44 documents across all 11 form types (Table 1).

4. Findings

We organize our findings around three themes, each corresponding to one of the research questions in Sec. 1. Section 4.1 reports how Care Fish was integrated into documentation work, speaking to the engineering question of whether it could be implemented as a working tool. Section 4.2 reports where it met friction with existing practices, including incompatibility with other information systems and differences from workers' individual writing styles. Section 4.3 reports changes in what was recorded, how different forms of evidence were valued, and how recording shaped the case interview itself; this material forms the basis for the social-science and design questions taken up in Sec. 5.

4.1. Making the Invisible Visible

4.1.1. Surfacing missed information.

Participants reported that Care Fish reduced documentation time by organizing conversation content that would otherwise require re-listening to long recordings. P2 noted that unlike voice recording alone, where he would need to search for relevant segments himself, Care Fish "*organizes it once, so time is definitely saved and I miss fewer things.*" For P3, Care Fish's annotation-style display revealed parts of the conversation he had overlooked: "*Because it shows the basis for its judgment as an annotation next to it, I can notice parts I had missed when looking only at the raw data.*" Care Fish organized the client's life story chronologically, surfacing stories that P3 had missed.

4.1.2. Worker agency over AI outputs.

All participants viewed themselves as able to at least partially control Care Fish. They reported that its errors were visible, verifiable, and correctable. Participants used Care Fish's outputs as a reference, and they maintained accountability for the final result. P3 gave the example of a client

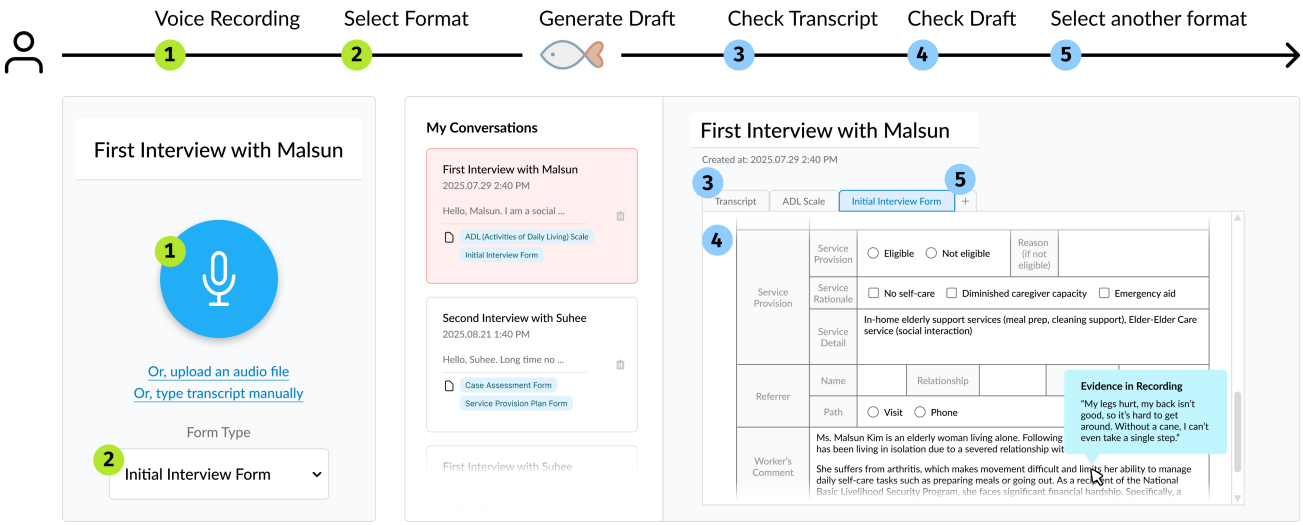

Figure 1. Use scenario of Care Fish. Users (1) record or upload audio from a case interview, (2) select the document format, and the system generates a draft. Users can then (3) review the transcript, (4) check the AI-generated draft with evidence from the recording, and (5) generate additional document types from the same recording.

with no history of cancer whose form was checked as having cancer, and whose gender was marked incorrectly: "*It's just obviously wrong, so it stands out.*" He noted that Care Fish "*doesn't get things wrong in ambiguous ways—it gets things clearly wrong,*" which made errors easy to catch during review. In terms of the first research question, workers treated Care Fish as a working tool that they, and not it, were accountable for.

4.2. Friction with Existing Practices

Care Fish entered a documentation practice that already had its own information systems, writing conventions, and workflows. Two forms of friction surfaced in our deployment.

4.2.1. Re-entry as accidental safeguard.

Care Fish was not compatible with the information systems workers already used. Participants re-entered Care Fish's outputs into other systems such as CareVisit or Excel. P2 described this as routine: "*We don't use it as-is; we go through the process of transferring it once more, and in that process, things get filtered out, so it wasn't a big problem.*" During one such transfer, he discovered that Care Fish had generated an incorrect service plan based on a misunderstanding of the conversation. The manual transfer forced workers to review each item.

4.2.2. Customization and consistency.

Social workers bring different writing conventions and documentation styles to their practice. P2 observed that "*even in official documents, some social workers use honorifics while others do not—everyone is still different.*" P3 noted that Care Fish was difficult to customize to these individual preferences, unlike ChatGPT where he could set prompts to match his style. He also valued Care Fish's consistency: "*Care Fish is more standardized. It's programmed, so when you input something, it comes out by form type.*" He contrasted this with ChatGPT, which he found "*unstable*" over time as its memory drifted from other uses. P3 also expressed concern that novice workers who rely heavily on Care Fish might "*end up copying without developing their own language or professional growth.*"

4.3. Recording, Evidence, and What Gets Said

4.3.1. Asymmetries in the authority of evidence.

Both workers and Care Fish were regarded as having errors and biases, but participants did not weigh the two kinds of error in the same way. Workers retained formal authority over Care Fish's judgments. In practice, however, the transcript excerpts it cited came to function as authoritative facts against which workers measured their own observations.

Workers did not extend Care Fish's authority to everything it produced: where they could directly observe, they treated their own observation as the source of truth, and where what the conversation meant was open to interpretation, they drew on their professional judgment to verify what Care Fish had produced. P4 described one example. She observed that Care Fish tended to assess conditions as more severe than workers would, particularly on evaluation forms involving clinical indicators, and she checked generated results against "*what I saw*" during home visits. P2 gave a different kind of example: Care Fish had added home healthcare to a client

service plan even though the client had clearly ruled it out for economic reasons. The AI had registered the recurrence of the term; P2, reading the same exchange, drew the opposite, and correct, inference: "*The AI just checked the box for home healthcare. Her wording afterwards was a bit vague, but she said things like 'can't, even if killed,' 'no money,' which together meant she couldn't use home healthcare. The AI, though, picked up that home healthcare kept being mentioned, and ended up adding a home healthcare service referral in the final service plan.*"

Yet participants also described Care Fish as a corrective for workers' own biases in moments when their judgments lacked clear evidence. P1 critically reflected the validity of her own non-verbal observation in this way once Care Fish offered an alternative form of evidence: "*So I'm not sure whether it's right to write things down to that extent (including her subjective opinion about the client's health status), or whether I should only write things grounded in actual evidence—like the way Care Fish does, based on real recorded audio of exchanges.*" She suspected that a "*biased view*" was encapsulated in her non-verbal observation of the client (e.g., a client's frailty, facial expression, or living conditions) and her own emotional response toward them. She described these observations as less authoritative than entries the system could ground in the recorded, and in this sense "*objective*," transcript. P4 made a parallel point about Care Fish's standing: she noted that it could remove the worker's "*cognitive error or distortion*," and in this framing the system helps produce "*objective documents*."

Workers thus held formal authority over both Care Fish's judgments and the final document, yet the evidentiary force of the transcripts Care Fish surfaced was at times treated not merely as more reliable than their own observations from the visit but as the standard against which those observations themselves were corrected.

4.3.2. Changes in what was recorded and how it was said.

Records in case documentation have always been kept selectively. Workers do not transcribe everything from an interview; they remove and abridge content that does not fit their information needs, and they rely on notes to hold the conversation together in the moment. P4 described her existing practice of writing things down during interviews in small pieces, both to capture information that came up and to find her way back when the conversation drifted: "*If I input things one by one, I also take notes, and notes on the main information that comes up in between. But if I try to record sequentially while doing it, there's no room in my mind to wait until the next piece of information comes up. So there's the process of coming back, and the process of trying to remember—that's why I check my notes.*"

Clients, too, present themselves selectively. Workers referred to certain client statements as "*lies.*" P1 described a situation where a client presented herself as having less family contact than she actually did; rather than confronting the client, P1 adjusted her prioritization of cases: "*I just think, oh, she's in touch with her family now, and move other people up in priority.*" She noted the difficulty of these judgments, given that "*all of our clients are economically struggling—they are all having a hard time, so the criteria for whom to support are ambiguous.*"

After Care Fish was introduced, workers described several changes in case interviews. P1 noted that recording helped her track changes across visits: "*If I don't have records and can't remember, I just keep thinking about old things and keep missing things. So recording has been better.*" She gave an example: a week earlier she had written that a client was "*disappointed with family*" and "*hadn't been in touch for over a year,*" but at the next visit she found herself writing that "*the daughter visited at the holiday two months ago.*" The successive entries contradicted each other, and she relied on the most recent one to make sense of the case. Both P2 and P4 reported that some clients were reluctant to be recorded or held back once recording began. P2 described a case in which a client kept saying "*do we really have to record?*" and he was unable to get consent that day. P4 noted that while clients had never objected to her taking notes, recording can be received differently: "*about one in ten might say they don't want to be counseled if I'm recording, or might close off a bit and be more reluctant to open up.*" P1 also described the opposite effect in other moments. Without the interruption of note-taking, conversation came more easily: "*that part starts to come up naturally, and because I'm so much younger than the older clients, they talk with me like they're talking to their grandchildren, and they bring up things they had never mentioned before.*"

Workers also adapted their own conversation strategies. P2 described his approach as trying to "*bring back the necessary content*" from client visits, framing the interview partly as an extraction task. P3 put it more directly: "*Before, the interview would end as an interview and we'd come back and organize. Now I think during the interview about how to produce the output—about running Care Fish or GPT.*"

5. Discussion

The deployment surfaced shifts in documentation practice that go beyond the system's intended function as a drafting aid. We discuss how Care Fish reorganizes the authority of different forms of evidence, how it reshapes the welfare regime that workers and clients enact in case interviews, and what design considerations follow from these shifts.

5.1. Reinforcing the Epistemological Hierarchy

The AI scribe reinforces an existing hierarchy in which institutional language outranks both workers' observations and clients' speech. It does so through one mechanism: the *materiality* of recordings. Workers' accounts of how they weighed the recording against their own observations point to a shift in how different kinds of evidence are valued in documentation.

Recordings appear to be neutral accounts of what happened because of three material properties: comprehensiveness, immutability, and replayability. Together, these properties give the recording the look of a complete and objective account, even though it captures only the verbal dimension and misses body language and the worker's embodied observation during a home visit. Care Fish's evidence feature reinforces this effect, since each AI suggestion is presented alongside a transcript citation that grounds it in the recording.

The recording's materiality confers authority on the AI scribe. Because the recording is comprehensive, immutable, and replayable, the entries Care Fish grounds in a transcript citation come to carry an evidentiary weight that workers' own observations do not. Before Care Fish, workers' observations, client statements, handwritten notes, and professional judgment all coexisted as sources of documentation, with no single source carrying decisive weight. The deployment shifted this balance. Entries that could be grounded in a transcript citation were treated, by workers themselves, as standing on firmer ground than entries that rested on observation alone, as P1's uncertainty about whether to write based on what she felt or to write "like the way Care Fish does, based on real recorded audio of exchanges" makes audible.

We call this *the reinforcement of an epistemological hierarchy*. This is a structure in which institutional ways of knowing, expressed through standardized assessments, eligibility criteria, and formal categories, are positioned above the client's way of knowing through lived experience and personal narrative, and above the worker's embodied knowledge of what the home looked like. The hierarchy itself is not new; documentation has always involved translating complex lives into institutional categories. What changes with an AI scribe is that this translation gains a material infrastructure (recordings that can be replayed, citations that can be checked) that makes it appear less like an interpretation and more like a neutral transcription of facts. As Dourish (2017) reminds us, the materiality of digital information is never inert; it shapes what counts as evidence and how evidence circulates.

This hierarchy raises a further question: *whose* way of knowing, exactly, sits at the top? The 'institutional categories and formal templates' that gain authority here are not anonymous. They belong to a particular regime: one with its own criteria of recognition, its own grammar of need, and its own street-level enactment. We turn to that regime, and to its situated vision, in the next section.

5.2. The Situated Vision of the AI Scribe

The same materiality that confers authority in Sec. 5.1 also gives that authority a particular shape. The vision that the AI scribe reinforces is not objective, but *situated*: it is the vision of a welfare regime, made operational through the materiality of recording. Following Haraway, every claim to a 'view from nowhere' is in fact a view from somewhere, with a knower and a stake (Haraway, 1988). By *welfare regime* we mean not the aggregate of welfare policies but the governing order through which an institution decides (i) whose suffering becomes legible, (ii) which forms of need qualify for which resources, and (iii) what vocabulary a person must use to be recognized as a recipient.

A welfare regime is enacted in encounters, not in policy documents. What workers called client "lies," such as understating family contact, presenting oneself as more isolated than one is, are acts of translation into the grammar the regime can hear. P3's account of thinking during the interview about "how to produce the output" shows the regime working through the worker as well.

Care Fish does not merely accelerate this regime; it changes how it operates in practice. The regime's grammar, once filtered through the worker's memory and handwritten notes, now has a recording, a transcript, and a citation as its substrate, with consequences for both workers' discretion and clients' speech.

Workers' discretion becomes costlier to exercise. Social workers are paradigmatic street-level bureaucrats (Lipsky, 1980): they implement institutional rules while exercising discretion to adapt them to individual client situations. In our setting, what workers called client "lies" can be read in this light. When eligibility criteria are stringent and the line between who qualifies and who does not is ambiguous, strategic self-presentation by clients—understating family contact, preserving dignity in the face of dependency, managing the relationship with the worker—is part of how the boundary between the institution and the person is negotiated in practice. P1's description of how she handled a client whose account of family contact had shifted, by quietly adjusting her prioritization rather than confronting the discrepancy, illustrates the discretion workers rely on. If the formal criteria of the "deserving poor" are applied too strictly, some clients who need services lose them.

Care Fish makes the act of discretion visible, and therefore costly. Before Care Fish, a worker could choose not to

document certain information simply by not writing it down. With Care Fish, that information may already appear in a generated draft, backed by a transcript citation. The worker must now actively delete or override it, in a visible act that itself becomes part of the documentation trail. The re-entry practice we observed acted as one safeguard, forcing manual review. But the direction of integration in commercial AI scribes points toward tighter coupling between recording and system of record, which would remove this checkpoint.

The third dimension of the welfare regime, namely the vocabulary one must speak in to be recognized, is also reshaped by recording. One worker anticipated that recording could make some clients reluctant to open up, while another described that clients spoke more freely because note-taking no longer interrupted the conversation. In both cases, the interview has become an event whose verbal content will be systematically retained and later processed. What used to be a conversation that a worker would selectively remember is now a recording that an AI will exhaustively read. P1's account of two contradicting entries, "disappointed with family" one week and "daughter visited at the holiday" the next, illustrates the shift from the worker's side: the client's own account of family contact, which used to live between visits and fade with the worker's memory, is now pinned down in the record and available for later comparison. Workers' own conversation strategies shifted accordingly, as P3's remark about thinking "during the interview about how to produce the output" makes explicit. The interview has become both a vehicle for assessment and the input to a documentation system, and participants on both sides of it act with that in mind.

This pattern connects to a broader observation in scholarship on public institutions: technologies that promise efficiency and transparency can subject vulnerable populations to a level of scrutiny that amounts to surveillance (Eubanks, 2018). AI scribes in social work sharpen this dynamic. The situated vision of an AI scribe is asymmetrically distributed: an exhaustive view of the client, paired with an unexamined view of the institution that judges them.

5.3. Design Considerations

The asymmetry described in Sec. 5.2, an exhaustive view of the client paired with an unexamined view of the institution, is the design problem this section addresses. If documentation AI systems embody a particular perspective through the templates they fill and the evidence they cite, what design strategies might help workers recognize and negotiate with that perspective? We raise three directions.

The first concerns **omission visibility**: current AI scribes show what they included, but not what they left out. The template-driven nature of Care Fish means that conversation segments not mapping onto any form field disappear without record. Making such omissions visible would allow workers to assess whether client concerns are being systematically overlooked by the template structure itself.

The second concerns **source attribution**: annotating each document entry with whether it derives from the client's direct statement, the worker's observation, or an AI inference from institutional criteria. Such markers would make explicit that different entries rest on different kinds of authority, returning to the worker the question of how to weigh transcript-grounded entries against observation from the home visit.

The third concerns **institutional criteria visibility**: rather than surfacing eligibility-relevant information about the client to the worker, the system would display which eligibility rule and which form template are driving each generated entry. What is made accountable shifts from the worker's exercise of discretion to the institution's grammar of recognition.

Each direction carries its own risks. Omission visibility could overwhelm workers with information. Source attribution could create a false sense of epistemic certainty. Surfacing institutional criteria could make eligibility rules feel more fixed than they are.

A related question concerns how much individual customization documentation AI systems should support. Workers asked for the system to accommodate their writing conventions, while at the same time valuing its consistency and worrying that novice workers might copy its outputs in place of developing their own voice. Granting wide latitude for per-worker customization would bring a documentation AI closer to a general-purpose writing assistant and would embed each worker's framings into the output in ways that may be hard to audit across cases. With a constrained system like the probe we deployed, the review that re-entry forced was one reason workers felt the final document remained theirs.

6. Limitations and Future Work

Our participant sample (N=4, with one in a lab setting) limits generalizability, though this is consistent with the technology probe methodology, which prioritizes situated understanding over statistical power. We thus do not claim generalizability or statistical validity for our findings; what we claim is the value of the critical questions they raise about AI scribes in social work, particularly how such tools may reinforce epistemological hierarchies. This aligns with the aims of big-Q qualitative research (Kidder and Fine, 1987). Our findings are situated in the Korean in-home older adult care context, shaped by the four participating workers, their institutions, and our own positionality as researchers; the dynamics of authority, discretion, and self-presentation may

operate differently in other social work systems. Future work should deploy AI scribes with more social workers across different institutional contexts.

Another limitation is the absence of the clients' perspective. We did not interview the clients, the older adults receiving care, which prevented us from directly analyzing behaviors such as refusal and deliberate self-presentation. Future work could engage older adults not as passive subjects of the technology but as participants in its design (Lazar et al., 2025).

We also did not collect fine-grained usage logs, so our evidence on usage remains anecdotal. Future work could log usage systematically, in particular recording both the frequency of re-entry catches and the type of error each catch surfaces, since the discussion in Sec. 5.2 treats re-entry as the main checkpoint that survives in current workflows. Finally, we intentionally left the design considerations conceptual to keep the solution space open; future work should develop and evaluate concrete implementations of them.

7. Conclusion

We presented Care Fish, an AI scribe technology probe for social work documentation, and reported findings from a field deployment with in-home older adult care workers. Care Fish reduced documentation time and surfaced information that workers had missed during case interviews, and its errors were visible enough that workers corrected them during a re-entry step into their existing systems. The deployment also surfaced shifts that went beyond usability: recordings and AI-generated citations acquired a form of authority that workers compared against their own observations, and the expectation of being recorded changed how both clients and workers spoke during interviews. We discussed how these shifts bear on workers' professional discretion and clients' self-presentation, and identified design considerations that respond to the asymmetry between the system's exhaustive view of the client and its unexamined view of the institution that judges them.

Acknowledgements

This study was supported by the Soldiers' Value and Culture Research Center (26-SVCR-02) of the Korea Military Academy.

References

Ali Alkhatib and Michael Bernstein. Street-level algorithms: A theory at the gaps between policy and decisions. 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. URL https://doi.org/10.1145/3290605.3300760.

David Y. Cao, Jamie R. Silkey, Michael C. Decker, and Karolyn A. Wanat. Artificial intelligence-driven digital scribes in clinical documentation: Pilot study assessing the impact on dermatologist workflow and patient encounters. *JAAD International*, 15:149–151, 2024. ISSN 2666-3287. URL https://doi.org/10.1016/j.jdin.2024.02.009.

Paul Dourish. *The Stuff of Bits: An Essay on the Materialities of Information*. MIT press, 2017.

Virginia Eubanks. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press, 2018.

Apoorva Gondimalla, Varshinee Sreekanth, Govind Joshi, Whitney Nelson, Eunsol Choi, Stephen C. Slota, Sherri R. Greenberg, Kenneth R. Fleischmann, and Min Kyung Lee. Aligning data with the goals of an organization and its workers: Designing data labeling for social service case notes. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703300. URL https://doi.org/10.1145/3613904.3642014.

Donna Haraway. Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, 14(3):575–599, 1988. ISSN 00463663. URL http://www.jstor.org/stable/3178066.

Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B. Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, Nicolas Roussel, and Björn Eiderbäck. Technology probes: Inspiring design for and with families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 17–24, New York, NY, USA, 2003. Association for Computing Machinery. ISBN 1581136307. URL https://doi.org/10.1145/642611.642616.

Eunkyung Jo, Young-Ho Kim, Sang-Houn Ok, and Daniel A. Epstein. Understanding public agencies' expectations and realities of ai-driven chatbots for public health monitoring. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400713941. URL https://doi.org/10.1145/3706598.3713593.

Louise H. Kidder and Michelle Fine. Qualitative and quantitative methods: When stories converge. *New Directions for Program Evaluation*, 1987(35):57–75, 1987. URL https://doi.org/10.1002/ev.1459.

Tuija Kuorikoski. Client documents in social work with adults as research data: Scoping review of opportunities and challenges. *Nordic Social Work Research*, 14 (4):532–544, 2024. URL https://doi.org/10.1080/2156857X.2022.2130406.

Amanda Lazar, Robin N. Brewer, and Bran Knowles. HCI and older adults: The critical turn and what comes next. *Foundations and Trends in Human-Computer Interaction*, 19(2):112–212, 06 2025. ISSN 1551-3955. URL https://doi.org/10.1561/1100000094.

Craig Lee, Shawn Britto, and Khaled Diwan. Evaluating the impact of artificial intelligence (AI) on clinical documentation efficiency and accuracy across clinical settings: A scoping review. *Cureus*, 16(11), 2024. doi: 10.7759/cureus.73994.

Hee Soung Lee and Soon-Ho Kwon. Problems and suggestions of welfare system for the elderly in super-aged society. *The Journal of Labor Law*, 50:1–29, 2020. ISSN 1229-4314. (In Korean).

Brenna Li, Noah Crampton, Thomas Yeates, Yu Xia, Xirong Tian, and Khai Truong. Automating clinical documentation with digital scribes: Understanding the impact on physicians. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. URL https://doi.org/10.1145/3411764.3445172.

Michael Lipsky. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. Russell sage foundation, 1980.

Jed Meers, Eppie Leishman, Izzie Salter, Simon Halliday, and Joe Tomlinson. What do the public think about artificial intelligence note-taking tools in social care? *European Social Work Research*, pages 1–16, 2026. doi: 10.1332/27551768Y2025D000000067. URL https://bristoluniversitypressdigital.com/view/journals/eswr/aop/article-10.1332-27551768Y2025D000000067/article-10.1332-27551768Y2025D000000067.xml.

Rajesh Nair, Muhammad Moinuddin Hashmi, Sameer S. Kassim, and Alexander Singer. The current state of digital scribes in primary care: A scoping review. *Journal of Medical Systems*, 50(1):2, 2026. ISSN 1573-689X. URL https://doi.org/10.1007/s10916-025-02319-4.

Whitney Nelson, Min Kyung Lee, Eunsol Choi, and Victor Wang. Designing LLM-based support for homelessness caseworkers. In *AAAI-2024 Workshop on Public Sector LLMs: Algorithmic and Sociotechnical Design*, 2024. URL https://openreview.net/forum?id=XyLgLwT5wJ.

Thu Nguyen, Alessandro Canossa, and Jichen Zhu. How human-centered explainable ai interface are designed and evaluated: A systematic survey, 2024. URL https://arxiv.org/abs/2403.14496.

Frederic G. Reamer. Documentation in social work: Evolving ethical and risk-management standards. *Social Work*, 50(4):325–334, 10 2005. ISSN 0037-8046. doi: 10.1093/sw/50.4.325.

Maxime Sasseville, Farzaneh Yousefi, Steven Ouellet, Florian Naye, Théo Stefan, Valérie Carnovale, Frédéric Bergeron, Linda Ling, Bobby Gheorghiu, Simon Hagens, Samuel Gareau-Lajoie, and Annie LeBlanc. The impact of AI scribes on streamlining clinical documentation: A systematic review. *Healthcare*, 13(12), 2025. ISSN 2227-9032. doi: 10.3390/healthcare13121447. URL https://www.mdpi.com/2227-9032/13/12/1447.

Cella M Sum, Anh-Ton Tran, Jessica Lin, Rachel Kuo, Cynthia L Bennett, Christina Harrington, and Sarah E Fox. Translation as (re)mediation: How ethnic community-based organizations negotiate legitimacy. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450394215. URL https://doi.org/10.1145/3544548.3581280.

Aaron A. Tierney, Gregg Gayre, Brian Hoberman, Britt Mattern, Manuel Ballesca, Patricia Kipnis, Vincent Liu, and Kristine Lee. Ambient artificial intelligence scribes to alleviate the burden of clinical documentation. *NEJM Catalyst*, 5(3):CAT.23.0404, 2024. URL https://catalyst.nejm.org/doi/full/10.1056/CAT.23.0404.

Amy Voida, Ellie Harmon, and Ban Al-Ani. Homebrew databases: Complexities of everyday information management in nonprofit organizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 915–924, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450302289. URL https://doi.org/10.1145/1978942.1979078.

Heesook Yoon and Mingyu Yoo. A study on improving the operation of Seoul's in-home care support centers for the elderly in 2023. Research report, Seoul Welfare Foundation, Seoul, South Korea, 2023. URL https://wish.welfare.seoul.kr/swflmsfront/board/boardr.do?bmno=10015&opno=10005&bno=100105. (In Korean).