Explainable Notes: Examining How to Unlock Meaning in Medical Notes with Interactivity and Artificial Intelligence

Hita Kambhamettu
hitakam@seas.upenn.edu
University of Pennsylvania
Philadelphia, PA, USA

Danaë Metaxa
metaxa@seas.upenn.edu
University of Pennsylvania
Philadelphia, PA, USA

Kevin B. Johnson
kevin.johnson1@pennmedicine.upenn.edu
University of Pennsylvania
Philadelphia, PA, USA

Andrew Head
head@seas.upenn.edu
University of Pennsylvania
Philadelphia, PA, USA

ABSTRACT

Medical progress notes have recently become available to patients at an unprecedented scale. Progress notes offer patients insight into their care that they cannot find elsewhere. That said, reading a note requires patients to contend with the language, unspoken assumptions, and clutter common to clinical documentation. As the health system reinvents many of its interfaces to incorporate AI assistance, this paper examines what intelligent interfaces could do to help patients read their progress notes. In a qualitative study, we examine the needs of patients as they read a progress note. We then formulate a vision for the *explainable note*, an augmented progress note that provides support for directing attention, phrase-level understanding, and tracing lines of reasoning. This vision manifests in a set of patient-inspired opportunities for advancing intelligent interfaces for writing and reading progress notes.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in interaction design**.

KEYWORDS

augmented medical texts, progress notes, patient-provider communication, intelligent reading and writing, attention, phrase-level understanding, lines of reasoning

1 INTRODUCTION

In recent years, patients have been given unprecedented access to highly-technical, detailed information about their care. In 2016, the 21st Century Cures Act was signed, guaranteeing patients in the United States access to their medical records [29]. Among these

records are a patient's "progress notes," or the notes their physicians and nurses (whom we refer to as clinicians) take during a visit. Progress notes detail the reasons for a visit, the tests conducted, and the clinicians' conclusions and care recommendations. The advantages of access to such notes is well-studied: patients with access to their progress notes report greater agency in their care [52], notice discrepancies in their records [23, 41], better remember the medical guidance they have received [18], and help caregivers understand critical information about patient care [12]. In short, patients benefit from reviewing progress notes.

While there are advantages to reading one's progress notes, ordinary patients are sometimes unable to understand their intricacies [60]. Currently, progress notes are largely written by clinicians for themselves, other clinicians, billing teams, insurance, and other stakeholders—but not patients [1]. These notes remain full of specialized terminology that is part of the medical practice [17, 65], and they leave unspoken the meaning of medical observations [17, 21] for which the interpretation is often self-evident to care teams.

In an era where artificially intelligent systems are increasingly being used in health care, this paper explores the role that intelligent interfaces should play in helping patients get the most out of their progress notes. Galvanized by advances in NLP technology, the HCI community has recently posited the roles intelligent UIs can play in supporting reading by explaining terms [5, 26], answering questions [5, 30, 85], assisting skimming [5, 20, 42], and expanding and compressing texts [27], among other affordances.

This paper examines what needs patients have when reading progress notes, and asks what it would mean to address them in the intelligent interfaces used for reading and writing notes. To answer this question, we conduct a patient-centric qualitative study. We opted for a methodology that combined observation, interviewing, and design feedback to come closer with the realities of reading. We conducted a 15-patient study wherein patients were asked to read through a recent progress note from their own care history and to speak aloud about the aspects of the text that they found difficult to engage with. Following the reading task, we spoke with patients about how to improve the reading experience, introducing mockups as probes to encourage reflection on concrete designs.

Our analysis of study sessions characterized three needs that arose during reading. The first was *directing attention*, where the information patients desired was often nested within an abundance

of irrelevant information. The second was *phrase-level understanding*, where patients had trouble understanding the terminology in a note, and found some terms alarming. The third was *tracing lines of reasoning*, where patients desired to understand how the clinician arrived at a health assessment and what the clinician believed the implications of lab and test results were.

Drawing together findings from the study, we articulate a vision of the *explainable note*: an intelligent, interactive note that provides integrated support for directing attention, phrase-level understanding, and tracing lines of reasoning, on the basis of a patient's goals and knowledge. We base our vision in a set of opportunities for augmenting note reading and writing interfaces that arose during conversations with patients. We describe these opportunities in detail, amidst considerations for preserving the original note's content and responsible integration of AI-generated content.

Altogether, this paper contributes a vision of the explainable note, and opportunities for intelligent interfaces to support its creation and reading, grounded in a qualitative study with patients.

2 BACKGROUND AND RELATED WORK

We begin this discussion with a primer on progress notes and their contents. Then we discuss what is known about the experience of reading notes. We then situate our findings amidst recent HCI discourse about the affordances that reading tools should provide generally and for clinical texts specifically.

2.1 Progress notes

A progress note is a healthcare document written by a provider to record what took place during a visit or encounter with a patient. Notes are typically written during a medical visit, and are often edited after the visit. The contents of notes often adhere to the SOAP format [76]: they contain *subjective* (S) observations of a patient's condition such as their spoken description of the symptoms they present with; *objective* (O) measurements like vitals and lab results; *assessments* (A) of the patient's condition, including diagnoses and prognoses; and *plans* (P) for how the patient's care should proceed, including prescriptions and plans of upcoming treatment.

Historically, progress notes have been written by clinicians to be read by other clinicians, for the purpose of facilitating continuity of care [1]. That said, for decades there have been calls among patient advocacy groups, like the Open Notes movement [15], to grant patients access to their own notes. Studies have shown myriad benefits of reading one's notes. Patients who read their notes report being able to better recall the details of their visits [18, 22], better understand their clinician's thought process [25], feel greater agency over their care [18, 52], and have greater confidence in the care they receive [22]. In some cases, patients have been able to identify inconsistencies in their record or errors in their care [23, 41]. There is evidence that sharing notes with patients improves both quality and safety of care [36, 77]. Patients and clinicians alike have reported that care would be improved if patients have access to their health information [22, 43].

For years, access to progress notes in the United States was limited to healthcare systems that opted to share them with patients. That changed in 2016 with the passing of the U.S. 21st Century Cures Act, which required healthcare organizations to give patients access

to their records, including progress notes [29]. Currently, patients in all U.S. health systems are able to access their progress notes—either by requesting them from their clinicians, or more commonly by accessing them through online patient portals. The recent increased availability of progress notes has spurred research interest in the informatics of these notes (see, for instance, [8, 9, 43, 54, 59]).

2.2 The note reading experience for patients

Prior research has revealed some issues that arise when patients read progress notes. For instance, patients have reported having trouble understanding jargon in their medical records [17, 65]. (And, it should be noted, even if a patient believes they understand medical terminology, they might still harbor misconceptions about that terminology [70].) Patients desire help in interpreting quantitative data in their record [21], such as labs and other diagnostic tests [17]. Patients have also requested help in identifying parts of notes that can help them understand what has changed in their care [17].

The purpose of our study is to deepen our understanding of the issues that arise while patients read their progress notes and outline potential solutions. We characterize the above issues into a set of three needs: directing attention, phrase-level understanding, and tracing lines of reasoning. In Section 8.1, we crystallize where our findings extend prior knowledge.

It bears mentioning that the level of confusion patients experience when reading notes seems to vary between patients. Some studies suggest that the broader population of patients as a whole do not find notes confusing, at least not to the extent of preventing them from learning something useful from their notes [60, 64]. It seems that patients who have more experience reading notes report less confusion when reading them [18]. In one study, patients who were older, less educated, unemployed or retired, or had lower levels of self-reported health reported greater difficulty reading notes [60]. Our study samples heavily from two of these segments, namely older adults who are retired. While patients in our study largely reported not finding notes difficult to read, nearly all demonstrated readability issues during the read-through activity.

The literature documents other points of friction in reading notes. Sometimes, medical notes contain mistakes [8] and inconsistencies [38]. In one study, patients reported 40% of these mistakes as serious [8]. Notes sometimes contain stigmatizing language of the kind that might alienate a patient [56]. Furthermore, some patients report worrying or becoming upset as a result of reading their notes [6, 60, 61]. Our study discusses the latter issue as it relates to understanding phrases in notes; we leave the other issues for deeper consideration in future research.

2.3 Intelligent UIs for reading progress notes

The focus of this paper is on understanding how patients can be aided in reading their notes with intelligent interfaces that support their creation and reading. The HCI literature offers inspiration for what affordances such interfaces might have. This paper considers these affordances as a starting point for investigation.

Recent HCI research has investigated challenges that arise during reading tasks with the intent of motivating the design of intelligent reading interfaces. Perhaps most relevant to our work, August et al. [5] observed healthcare consumers as they read medical research

articles. They observed that participants had trouble understanding jargon, hindering their ability to read passages filled with such terms. Participants were not sure what to read in the article, or where to find information that they knew they wished to find. We suspected many of these challenges had analogs for readers of medical notes, and in fact our study showed these challenges and others manifesting in particular ways in notes. Other studies outside of the medical domain have shown readers having trouble understanding terms [26] and hidden details [55]. Additionally, readers have questions about facts, reasoning about document contents, obtaining document overviews, and accessing external resources while perusing documents [30]. These studies have spurred systems development in intelligent interfaces to address those issues, and we hope our study will do the same in the domain of medical notes.

We draw inspiration from research that has proposed how intelligent interfaces can help people read difficult texts outside of the setting of medical notes. A recent useful framing is provided by the Accessible Text Framework [27], which proposes that reading interfaces be augmented with new ways to compress, expand, experience, and review text. Among the proposed affordances are the ability to summarize and prioritize passages, reduce reading volume, apply lexical simplifications, and explain context relating to a passage. Many of these affordances map to the needs our study characterizes and the opportunities for design we propose. Where our paper goes further is in describing what such affordances would need to look like in the context of medical progress notes. Specifically, we detail for what passages they would be activated, and delve into the intricacies of what it would take to provide meaningful support for reading progress notes.

Myriad recent HCI systems have embodied affordances that, if adequately tailored, could support the reading of progress notes. These systems have provided support for readers to access definitions of terms and symbols [26], skim documents [5, 20, 42], navigate them [5], annotate passages [74], take and manage notes across documents [34, 35], access supplemental material related to a passage [13, 32, 37, 57], preview the contents of cited material [57], and get answers to their questions within-document [30, 85]. Many of these affordances could support the reading of progress notes.

Patients might also be better supported in learning their notes if they had better tools to understand the health data within them. Health data has been characterized as difficult for patients to understand [19]. Interfaces have been proposed to scaffold understanding of health data by providing deeper connections from natural language descriptions to elements of health visualizations [75], supporting collaborative reflection on data with healthcare professionals [48], and making the causes of symptoms more visible to a patient [68]. Interactions like these could serve as interaction primitives for understanding data in progress notes, if the data is sufficiently important and prohibitively complex.

The purpose of this paper is to highlight the medical note as an apt site of augmented reading and sensemaking. Furthermore, we elucidate what it means to tailor known HCI interactions to the note, answering questions like what content should be emphasized, what good definitions might look like, and what kinds of questions readers might need to have answered. We highlight our key takeaways for design in Section 8.1.

A related line of work has explored how interfaces might help clinicians read clinical notes. This work is in part motivated by the considerable amount of time clinicians spend reading clinical notes (by one estimate, around 15% of a clinician's working hours [2]). Notes have been described as causing issues of information overload and underload (i.e., the absence of needed information) [7]. An additional challenge is that notes contain a great deal of duplicated and templated text [62, 81] that clinicians must sift through.

To address these challenges, the HCI and medical communities have proposed interfaces that help clinicians read notes, many of which have affordances that could map nicely to the situation of patients reading as well. For instance, these systems help clinicians disambiguate medical acronyms [53], look up vitals or lab results at the site where it is described in the note [53, 80], highlight passages of a note related to a medical concept of interest [67, 71, 73], navigate between semantically-related passages in consecutive notes [71], review how test results have changed over time [45, 72], review automated medical recommendations [67], pull up decision support tables [67], visually assess which test results are out of range [45], retrieve relevant medical images [11], and skim [72]. Our study suggests that patients would appreciate similar features to some of these in their reading interfaces with appropriate adaptations, namely that they may need additional support for phrase understanding and interpretation. We outline those opportunities, emphasizing the patient's perspective throughout our characterization.

Prior work has also augmented clinicians' writing tools to assist in writing quality notes. These systems assist in the entry of repetitive text [53, 63], support annotation of the medical record [72, 78], and help clinicians embellish notes with graphics, handwritten notes, and transcripts of conversations with patients [78]. They have also supported the automatic generation of patient-friendly views of notes [67]. Our paper briefly discusses the role augmented writing tools could play in supporting the creation of notes that will be more readable to patients.

3 METHODS

To crystallize the vision of explainable notes, we conducted a qualitative study with interview, observation, and design feedback components. The study was designed to answer two research questions:

RQ1: What are the needs of patients when reading their medical progress notes?

RQ2: How could interfaces to progress notes be designed to address these needs?

To answer these questions, we designed the following study.

3.1 Participants

15 patients were recruited from a patient and family advisory council (PFAC) at the University of Pennsylvania Health System. The leader of the PFAC reached out to potential participants on our behalf. Those who were interested in participating reached out directly to our research team. Thereafter, we recruited additional participants through snowball sampling. As a result, 13 of 15 participants were PFAC members and 2 were not (though all receive care from the same healthcare system).

14 of 15 patients provided demographic information. Of these, 57% identified as female and the rest (43%) as male. Ages ranged between 23 and 84 years old, with a median age of 62. 78.6% reported their race as Caucasian/European/White, 14.3% as Black/African American, 14.3% as Asian, and 7.1% as another race or ethnicity. About half (7) self-described as retired, and 2 as semi-retired consultants; 2 described themselves as clinical research coordinators; 1 as an auditor; 1 as a housewife; 2 as a teachers/consultants; and 1 as self-employed. Our sample therefore heavily represents older adults and retirees. While it underrepresents other groups, we note that this group tends to make heavy use of the health system and for that reason may find particular value in access to notes.

Patients were managing a wide variety of conditions, within the specialties of pulmonology, cardiology, ophthomology, and lymphology, among others. Eight patients were managing chronic conditions. Patients were also asked to comment on their prior experiences with progress notes. First, they were asked how often they checked their progress notes. 20% had never checked their notes before, 53% checked their notes less than monthly, 7% checked their notes on a weekly basis, and 13% checked their notes on a daily basis. Altogether, patients reported a relatively high degree of understanding their notes: on a 5-point Likert scale (where 5 indicated strongly agreement that they understood the content of their clinician's notes), 33% reported a comprehension level of 5, 33% a level of 4, and 13% a level of 3 (the remaining 20% had not read their notes before). We note that while patients largely reported understanding their notes, nearly all pointed out passages that caused difficulty during a reading task. Individual patient profiles (including age, gender, ethnicity, kinds of medical conditions, and note reading experience) appear in Appendix Table 1.

3.2 Procedure

Prior to study activities, we undertook ethics review with our institution's IRB. Study sessions lasted for one hour each. Each session took place over Zoom, so patients could participate from a location that was comfortable to them. Study sessions lasted for one hour each. Each session consisted of three parts:

1. Briefing. We defined what a progress note was and told the patient the purpose of the study. Participants were asked for their consent to participate, and for their consent for us to record audio and screen recordings from the session.[1] They were told that the researchers were not affiliated with their hospital system or their medical providers. Then we asked the patient to describe why they read their progress notes (if they have done so before).

2. Observed reading task. To better understand the needs of patients when reading progress notes, we observed each patient as they read one of their recent progress notes. The patient logged into their patient portal and selected a recent progress note. They were asked to prioritize a note that represented a recent, significant visit. Then, they read the note. As they did so, we asked them to narrate which parts of the note contained useful information, and to describe the parts of the note that were particularly difficult to understand. We often asked follow up questions to better understand

the nature of the difficulties encountered. The patient shared their screen so they could point to specific passages they were discussing. Throughout this and the next stage of the study, we took the duty of care measure [3] of encouraging patients to speak with their healthcare providers if they had negative reactions to information in their note; we did this with two patients.

3. Interviews assisted by mockups. We then conducted a semi-structured interview with the patient to discuss how notes could be augmented to support their reading. We began our conversation by reviewing challenges that arose during the reading task. As we discussed each challenge, we also asked about solutions.

Patients sometimes volunteered their own ideas for solutions. To deepen our conversation of solutions, we sometimes showed patients mockups (i.e., still image prototypes) of augmented notes that we had created before hand, if they related directly to issues brought up in the conversation. When we showed a patient a mockup, we asked the patient if the mockup would have helped them in their reading, and what could make it more useful. To reduce anchoring bias, we only showed a particular mockup after a patient volunteered a relevant reading obstacle. Mockups were created using excerpts of progress notes from MIMIC-III [33], a publicly available clinical database comprising deidentified health records.

Mockups were developed for the following preliminary ideas:

- *summary*: an example jargon-heavy passage was augmented with an AI-generated summary paraphrasing the passage in more patient-friendly language.
- *diagnosis explanation*: a mention of a diagnosis was augmented with an AI-generated note in the margin that explained how a clinician arrived at a health assessment, written in patient-friendly language.
- *lab interpretation*: a lab result found in the note was augmented to highlight which of many lab results were pertinent to a patient's diagnosis, with AI-generated explanations describing the significance of values that fell outside of normal ranges.
- *testimonial*: a mention of a medical condition was augmented with a margin note showing excerpts of a patient-written testimonial (e.g., as if scraped from Reddit) describing a patient's experience with that condition.
- *messages of comfort*: a concerning term (an undesirable health condition) was augmented with a message from a clinician stating that the patient need not be worried about having this condition.

Collectively, these mockups were designed to address challenges we had identified from the literature and pilot interviews, including understanding jargon [17, 65] (*summary, testimonial*), navigating cluttered content [7] (*summary, lab interpretation*), *interpreting data* [17, 21] (*diagnosis explanation, lab interpretation, messages of comfort*), and interpreting alarming information [6, 60, 61] (*testimonial, messages of comfort*). Appendix A.2 shows images and explanations of each mockup. The correctness of information in mockups was verified by a member of the author team who is a board-certified physician. Following the conclusion of the study, patients were given $25 USD as compensation.

---

[1]Following the study, we additionally contacted and obtained consent from individual participants whose notes we wished to excerpt in this paper.

3.3 Analysis

Our study yielded four kinds of data: audio recordings, screen recordings, patient's responses to the background questionnaires, and the researchers' notes. We aimed to uncover the needs of patients during the note reading process and potential solutions. To do this, we conducted a thematic analysis on the researchers' notes and the audio transcripts, referencing screen recordings for additional context when needed [10, Chapter 5]. Initially, one author created a set of codes through an open coding pass. Then, a second author reviewed these initial codes, collaborating with the first to refine the codebook. This refinement process involved removing codes that did not significantly contribute to understanding patient needs or suggest improvements to note-reading technology, such as codes related to inaccuracies in notes. The final codebook contained nearly forty codes. For validation, we engaged in detailed discussions at every stage of analysis, prioritizing collaborative review of the coding for its accuracy and consistency. This approach was chosen over calculating inter-rater reliability (IRR) because the detailed nature of the codes and the need for context-aware discussions offered richer insights [50].

4 FINDINGS

In the next three sections, we report our findings. The first section sets the stage by clarifying the value of reading progress notes as our informants saw it. The second section characterizes three patient needs that arise while reading progress notes, and explores opportunities to address these needs. The third section offers considerations into how to responsibly augment notes to support patient reading. Section 8.1 clarifies that our findings deepen the understanding of patient needs and outline how to address them in relation to prior work. Our findings are supported with anecdotes, quotes, and images from patients' notes. When excerpts of notes are shown, they are de-identified. Patients are referred to with pseudonyms P1–15. Quotes were lightly edited for brevity and clarity.

5 PREAMBLE: WHY READ PROGRESS NOTES?

Patients commented on aspects of progress notes that they found valuable. We describe them here to highlight what kind of outcomes would be achieved better if patients receive adequate support for reading their notes.

Recalling details from a visit. For many patients, progress notes were valuable because they helped them recall details from their visits ($N = 5$). P15 described that when they read their notes, "…*I don't remember for sure some specific thing, and I go back to look to see if it's there.*" Sometimes, patients noted that they were not in a place during their medical encounter to be fully present with their clinician. For example, P12 described that notes were helpful for reviewing their clinicians' impressions and particulars of medical procedures for which they had been "drugged up."

Some patients described that their care was handled by multiple clinicians ($N = 2$); for one participant, their clinicians belonged to multiple distinct health systems. For these patients, health records provided a means to self-educate about one's care enough to convey what they had learned from one clinician to the other clinician.

Assessing common ground. Patients might not know whether they left a visit on the same page as their clinician. Patients described one benefit of progress notes as letting them check that what they shared was correctly interpreted, and that their takeaways from the visit matches those of their clinician ($N = 3$).

Learning the subtext of a visit. Notes were seen as helping patients learn about aspects of their care that did not come across during the visit. In some cases, notes could convey a clinician's honest opinion—what P2 referred to as their "*actual, unspoken opinion*"—which they felt clinicians may have chosen to hold back ($N = 2$). Notes could also convey the attention that had been given to the patient's care ($N = 2$), or could help a patient "*feel seen*" (P1). For P8, reading progress notes helped them understand the work their clinician was doing behind the scenes; their note showed some of the work that the clinician had done to prepare referrals, communicate with other clinicians, and order medications, among other tasks. P11 was reassured to see the efforts their clinician had taken to collaborate with other clinicians, and to see their clinician acknowledge their limits in their understanding of P11's symptoms.

6 NEEDS AND OPPORTUNITIES

The main outcome of our study is a characterization of three needs that arise when patients read progress notes, and detailed observations of how they could be effectively addressed. Each need below is described in terms of a *problem* and the *opportunities* it entails. In general, the observations we share in the problem sections follow from patients' thinkaloud reports, reflections on the reading task, and reflections on past experiences with notes. Observations in the opportunities sections generally arose from discussion of the mockups. We explicitly state in the text whenever this is not the case. Below, we describe the three needs:

6.1 Directing Attention

The problem. Notes were written in a way that obfuscated information that patients cared about. Patients described their notes as a "*dump*" (P1), "*pages of kind of mush*" (P2), and a "*long block of text*" that "*bogs you down*" (P4). For many patients, the information they found most informative related to the clinician's assessments of their condition, and plans for their upcoming care ($N = 3$). This information was often buried in the note, appearing at the very bottom as might happen if a note is ordered according to "SOAP" convention (see Section 2.1). Patients therefore sometimes scrolled through considerable amounts of irrelevant information before finding information that was useful to them. P10 pointed proposed an alternative organization:

> I would almost rather see the assessment and plan at the top, because why should I have to scroll down to the bottom to figure out what I need to be doing?

Many patients pointed out parts of their notes that felt irrelevant ($N = 4$). Their notes appeared like the clinician was trying to "*write down everything*" (P9). Sometimes information was irrelevant because it was simply already known the patient. As P9 put it:

> I don't need to know what my prescriptions are, I know that already. It's a lot of information, and she's

simply writing down what I told her. I don't want to be reminded of everything.

Sections of the note that some patients found irrelevant included the past medical history ($N = 4$) and medication lists ($N = 2$). One such medical history section is shown in Figure 1; passages like this could be dense and long. Patients frequently believed that parts of their notes were copied from other places in their record ($N = 3$), or from different sections of the same note ($N = 1$). Sometimes, the included information felt extremely dated. P14 described a case where their clinician referenced a finding from about 60 years prior and asked "*who cares now?*" P10's note included pages of pasted text about their pancreatic cancer history copied from prior visits that spanned from weeks ago through about 8 years prior.

Patients also described skimming their notes ($N = 2$); P4 felt that clutter led them to skim ($N = 1$). Additionally, patients shared concern that they might miss something important if they did skim ($N = 2$). P13 for instance described that "*I might miss stuff because I'm thinking there's nothing else there but really they hid a little special thing... like the first [note I read], where they had upped [the dosage] from 2,000 a day to 4,000 a day.*"

Opportunity: Automatically emphasize important content. One role that intelligent interfaces could play in helping a patient read notes is in directing attention to the content that will be important to a patient. As P15 put it, the patient could receive help "*differentiating what's relevant and what's not, so that I know I can kind of skip unless I want to review where we've been and where we're going.*" One aspect of the mockups that resonated with patients was their use of standard visual emphasis primitives like bolding, highlighting, and font choice to make important text stand out ($N = 2$). Some of these affordances are already in notes—for instance, P10 pointed out that certain lab results were highlighted to indicate results that were out of range, and appreciated that "*I won't have to do a lot of work there because they highlighted the numbers that are different now.*" Patients desired similar emphasis of other kinds of information. P8 believed that the *lab interpretations* mockup (see Section 3.2) would "*answer a lot of questions*" highlighting the subset of tests that were relevant to a patient's diagnosis. P15 suggested that medication lists also be styled in a way to draw attention to those prescriptions that changed in dosage since prior visits.

Opportunity: Automatically deemphasize insignificant content. Patients also envisioned reading experiences where irrelevant parts of the note were deemphasized. As noted above, during the reading activity, patients often found sections like prior medical history and medication lists to be irrelevant. P9 called for such information to be removed from the note wholesale:

> Take out the medical history unless we discuss it. Take out the demographic information... I would expect the notes to be tailored to this exchange...they're not.

Other patients envisioned forms of deemphasis that preserved the original content. P2 asked for a note where they first saw high-level takeaways about the visit, and then could access details afterwards on-demand. P6 proposed initially hiding some classes of information by default and making them accessible by clicking simple labels, stating that "*allergies and medication you're on makes going through the note a little tedious. You can put in something like*

Past Medical History:
1) SSS (sick sinus syndrome)
2) VT (ventricular tachycardia)
3) Nonischemic cardiomyopathy
4) Ventricular tachycardia
5) Hyperlipidemia
6) Automatic implantable cardiac defribillator in situ
7) Non-ischemic cardiomyopathy
8) Cardiac sarcoidosis
9) HTN (hypertension)
10) Type II or unspecified type diabetes melititus without mention of complication, not stated as controlled
11) Elevated LFTs
12) Awaiting organ transplant
13) Squamous cell carcinoma in situ of skin
14) Malignant melanoma of skin

Figure 1: A passage that was seen as irrelevant by P14. In their words, "Most of these are not issues...because a lot of this stuff went away because of the transplant. Every time I see this past medical history it definitely makes me a little nuts because it's in the past... what about what's happening today?"

'click to see allergies' and 'click to see medications' and 'click to see past surgeries.'"

6.2 Phrase-level understanding

The problem. The most frequently reported impediment to reading was jargon. 14 of 15 participants pointed out jargon in their note that they found difficult to understand. Notes were described as using "*dense verbiage*" (P10) and that "*you have to be extraordinarily literate to get some sense out of this*" (P4). Jargon took the form of either medical words and phrases, or acronyms. 6 participants pointed out an acronym that was difficult to understand.

Understanding jargon was described as a "*constant headache*" (P10). One reason is that unfamiliar terms did not appear on their own, but rather as parts of passages with lots of jargon. Consider P11's experience when reading a passage that noted "*an isolated small ulcer at the ileocecal valve, and mild congested mucosa in the rectosigmoid colon with some cryptitis without changes of chronicity*":

> I don't know what an "ileocecal valve" is, I'll probably have to look that up. I don't really know what "cryptitis" is either, and at this point I'm just like whatever... I don't know what any of it means.

Consider also the jargon-heavy passages that P12 was reading, shown in Figure 2. Here, the patient reviewed the notes their clinician had taken on a pre-operative cardiac risk assessment that had been performed. The outcome of the assessment is made up of a number of observations, all written in specialized terms that refer to procedures (e.g., "lateral right tubutlar microdissectomy L5-2"), observations (e.g., "significant cardiac arrhythmias"), and measures (e.g., "RCRI"). The passage as a whole is difficult to understand, not to mention the specific phrase "dyspnea on exertion" that P12 had

Pre-operative cardiac risk assessment:
The patient is planning to undergo lateral right tubular microdisectomy L5-2 in the near future. He does not have any active cardiac conditions including unstable coronary syndromes, significant cardiac arrhythmias, decompensated heart failure, severe valvular stenosis. He does note dyspnea on exertion with <4 METS of activity so I recommended a pre-operative pharmacologic nuclear stress test. If the stress test is negative for ischemia, he may proceed with the planned procedure without additional testing. His anti-platelet therapy [aspirin] can be held for 5-7 days before his planned surgery/procedure and can be restarted the day after if there is no bleeding. See RCRI score below.

Figure 2: A passage from P12's progress note that makes heavy use of jargon. Medical terms and acronyms are highlighted in blue to emphasize the use of jargon; this single passage makes use at least a dozen medical terms.

singled out as something that they felt was important to understand. Some participants described giving up on trying to read a passage after encountering a lot of jargon ($N = 2$).

During our analysis, we believed that at least 3 participants misinterpreted the meanings of acronyms in their notes. For instance, P2 did not know the meaning of "ED" (which likely stood for "emergency department"). They hesitated to look up the meaning of the acronym on the web because they expected there would be multiple expansions of the acronym and no easy way to tell which one was correct. Eventually, they mistakenly concluded that the acronym might be related to the procedure they underwent during the visit.

Several patients described that they might look up the meaning of a phrase, for instance by using web search ($N = 3$), though we note that this may not be an adequate solution. As P2 pointed out, some medical terms have multiple meanings, of which only few are relevant to them. Some patients furthermore reported that they did not feel equipped to read the medical literature online ($N = 2$).

It was not uncommon for some of the words and phrases in a note to be worrisome ($N = 4$). P4 described how some phrases grab one's attention, sharing that when clinicians "*say a word like lesions, cancer… then it's like you can't ignore them.*" Phrases were described as setting off "*alarm bells*" (P9) or "*red lights*" (P10). Sometimes, it was exacerbated by the clinician styling text in a way that suggested concern, such as highlighting text in red (see Figure 3). P14 pointed out a passage where the clinician referred to them being at high risk for a condition referred to by an acronym:

> I see this sentence: 'The patient is at high risk for DM.' Well I don't know what that is, so yes it's concerning but it's not [supported with more details]. I would take this to my PCP and ask her if she's in agreement that I'm at a high risk for whatever this is. And if so, how come [she's] not telling me to [alter my behaviors or] calling the transplant team and telling them about my activities?

Phrases like these might be particularly useful to help patients understand the significance and implications of the medical assessments found in their progress notes.

Opportunity: Provide context-sensitive definitions of terms. Given how often patients spoke of jargon, it seems patients would benefit from easy access to definitions of unfamiliar terms and acronyms. Perhaps patients could be allowed to look up definitions right alongside the text. This was suggested by P5 after viewing the *diagnosis explanation* mockup when they asked, "*when there are acronyms like HTN, could I just highlight that and see it?*"

That said, there is considerable nuance in providing definitions in a way that is useful. P5 had clarified that sometimes medical definitions were not tailored to the patient; in the past they had looked up definitions and found it "*pretty useless because it was a blood test [and] they have all these different measures but they don't properly explain what it's measuring.*" If an acronym has multiple senses, as we observed sometimes happens above, the correct sense needs to be found. And furthermore, definitions should ideally be written in terms that match the context of the note. P15 described a prior experience reading a medical definition that was technically correct, but contextually inappropriate. They had looked up a medication their clinician had mentioned in their note called Atorvastatin in order to understand why it was prescribed. The definition told them the medication is typically used to mitigate high cholesterol, but because their visit was meant to address an eye infection, the medication appeared irrelevant. What the patient later learned was that cholesterol sometimes affects eye dryness. Without this context, the description of Atorvastatin was befuddling.

What is more, we note that not every patient requires help understanding the same terms. While many patients desired help understanding terms that were considerably specialized, others desired help for terms that have entered everyday lexicon. For instance, P15 told us that "*I even have to look up 'cataract' every time I see it because I use it so infrequently that I can't keep it in my head what it is.*" As a whole, it seems that for definitions to be useful, they should define terms unfamiliar to the reader, in a way that is in and of itself jargon-free, and sensitive to the context of the note.

Opportunity: Incorporate abstractive summaries. Another method to help patients cope with jargon is to eliminate it altogether by supplying patients with more readable summaries of their notes' contents. Some patients ($N = 3$) appreciated the plain language explanations in the *summary* mockup. P2 shared that such explanations are particularly helpful when the jargon-dense information appears alarming:

> By giving you more information in a readable format and not confusing you, it's gonna at least calm me down so I can wait a few days and hopefully get to the doctor by the end of the week…if you read an explanation in easier language to understand…it would be very helpful.

6.3 Tracing Lines of Reasoning

The problem. Patients often felt they did not understand the significance of information in their notes. They desired interpretation and contextualization of the kinds that could only come their
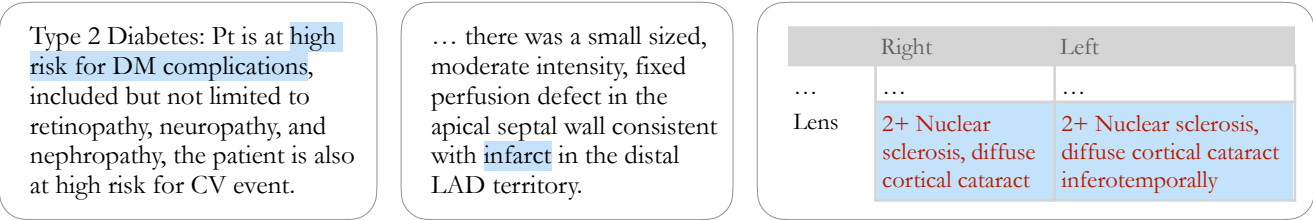
Type 2 Diabetes: Pt is at high risk for DM complications, included but not limited to retinopathy, neuropathy, and nephropathy, the patient is also at high risk for CV event.

… there was a small sized, moderate intensity, fixed perfusion defect in the apical septal wall consistent with infarct in the distal LAD territory.

		Right	Left
…	…	…	…
Lens		2+ Nuclear sclerosis, diffuse cortical cataract	2+ Nuclear sclerosis, diffuse cortical cataract inferotemporally

Figure 3: Many patients found terms that attracted their attention, sometimes in a negative way. Pictured are three passages from P12, P14, and P15. Phrases that attracted patients' attention highlighted in blue. The red color of the text in the last pane (for "2+ Nuclear sclerosis…") was added by the clinician, which is what triggered alarm for the patient.

clinicians, or others with shared health experiences. Most patients expressed a desire for more context from their clinician about information in their note ($N = 8$). This was described as understanding the thought processes ($N = 2$) of the clinicians, or "*getting in the doctor's head*" (P2). Some patients desired more insight into the next steps to take in their care ($N = 4$). Others wanted to know why a lab or test result was ordered ($N = 3$). Others wished for an understanding of how lab or test results related to a diagnosis ($N = 4$), perhaps in order to better understand that diagnosis ($N = 2$). Some examples of missing context and subsequent questions patients desired to be answered in their notes are described in Figure 4. P8 described frequently reading assessments in his notes to learn about their diagnosis of prostate cancer, and attempting to reconcile results with what it meant for the severity of their condition:

> [My clinician would write that] the next step is a biopsy. They didn't talk much about the Gleason scores or what they were. That was never communicated to me. I got that through digging through other treatment reports. I found out that the score generally gives you an idea of how aggressive the cancer is.

This experience was representative of several other patients, who also wished to understand the implications of a test result. P15 described finding a test result in their note that had been highlighted in the record—and so presumably had drawn the clinician's attention—but it had not been explained to them. In particular, a portion of the results for a slit lamp eye exam was highlighted in red, reading "2+ Nuclear sclerosis" (see Figure 3 right side):

> I want to know what [the clinician is] seeing and what it means. Is it something to watch out for? Is it important to get rid of it? What does it mean to your future, what is there to do about it?

Opportunity: Provide the missing interpretation. Perhaps notes could be augmented in a way that integrate the interpretation that patients desired into the note. P9 described one vision of what such interpretation to look like:

> This is what I want: 'I let the patient know she has this kind of cancer because her score is X. Her prognosis is Y. This is the treatment plan I'm recommending.' I want a translation of what's next for me.

In other words, the patient wanted a (notably brief) statement of the clinician's thought process, connecting a test result to a

prognosis and a set of next steps. Our *lab interpretation* mockup (Section 3.2) addressed this issue, and was seen as providing compatible information to what was desired. In the words of P8,

> [The design from this mockup] would really answer a lot of questions… it says what the recommendation is. I particularly like how the explanation explains why the clinician ordered the test, what it means given your past medical history it's linking back [to], and what the prognosis is and the next steps.

Opportunity: Incorporate messages of reassurance. Sometimes, notes lacked the interpretive layer necessary for patients to fully grasp whether certain phrases, which might appear alarming (see Section 6.2), were genuinely cause for concern. In these cases, notes could be augmented to provide assurance around alarming information. Patients seemed to desire more direct connection with their clinician within the notes. This was evident in their appreciation for a personal tone from clinicians, either observed in the actual notes or demonstrated in the mockups ($N = 3$). Specifically, P11 appreciated messages in the *reassurance* mockup that conveyed good bedside manner, aligned with the note's clinical content:

> I understand the need for clinical explanations in the assessment, but the gray box [in the mockup] adds that bedside manner… like 'this is what we really need to focus on' and 'this is what we're going to do,' 'I'm going to follow up with you,' and 'when I am going to do x, y, z…' I think that is very helpful. I love this.

Opportunity: Connect to second opinions. In some instances, patients felt that the interpretation most beneficial to them may not come from the authoring clinician, but rather specialists in other fields ($N = 3$). Managing one's care could require curating an understanding that spans multiple specialties. For instance, as noted in Section 5, patients like P6 have emphasized the importance of avoiding "*tunnel vision*" by their primary clinicians. P6's condition intersects multiple specialties, which made it necessary to "*acknowledge the other specialties' opinions to find the right plan for me.*" This multi-faceted nature of patient care suggests a potential improvement in how progress notes are constructed and used. A more holistic approach might involve augmenting notes to provide patients with broader, more contextual perspectives. P2 described this as "*getting around the things that [the clinicians] don't know*

HPI:
Patient feels the vision has been pretty much the same following the last exam. Although notices variable changes which tends to improve through blinking. Negative eye pain, discomfort, flashes, or floaters. Due to start Atorvastatin.

Biopsy from 12/24/20:
1. Right breast, core biopsy: Benign fatty breast tissue, no tumor seen.
2. Left breast, core biopsy: Infiltrating well differentiated duct carcinoma with focal lobular feature
3. HER2 oncogene expression: NEG (score 0)

TRANSTHORACIC ECHO (TTE) COMPLETE 6/2022
The left ventricle is normal in size. There is severely increased wall thickness consistent with severe concentric hypertrophy. LV IVS measures 1.8 cm. LV posterior wall measure 1.6 cm. There are no segmental wall motion abnormalities.

Q. *Is "negative eye pain" a good or bad thing?*

Q. *What prognosis is suggested by the score?*

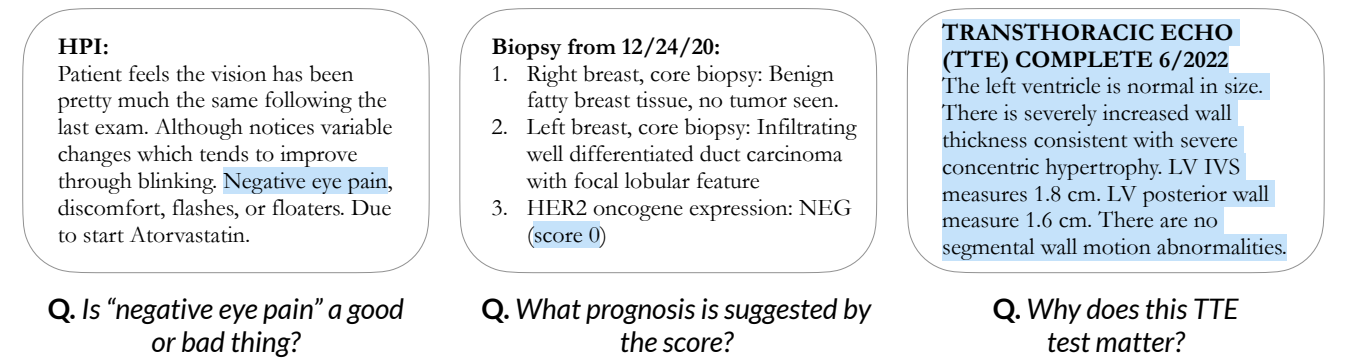**Q.** *Why does this TTE test matter?*

Figure 4: Patients read passages that they knew were related to their care, but they did not understand their significance. Pictured are passages from P15, P9, and P12, and the questions that they raised for each patient.

and understanding things in context." P5 suggested one solution: providing "*recommendations about following up on details, or consulting with other members of the medical team.*" By integrating advice on whom to consult or how to follow up, progress notes could become a more dynamic tool, offering guidance beyond the immediate clinical encounter, and in turn help patients trace lines of reasoning across a larger health system.

Opportunity: Relate to other patients' experiences. Patients saw value in the perspectives of other patients. This became apparent when they commented on the *testimonials* mockup. P8 articulated that seeing testimonials from patients with similar health backgrounds can expand one's view of available options for their care. They desired an understanding of "*all my options so I can decide what I like and what I don't like, what I don't understand, see what decisions I can make.*" P8 imagined it as being particularly useful to compare their condition's progression to others' and learning of the treatments others had undergone. P6 similarly saw the value of testimonials in the suggestions they might yield for better understanding or relief. P8, however, expressed a preference for keeping patient-sourced information distinct from the main body of the note to avoid "*muddying up*" the clinical content.

7 MINDFUL AUGMENTATION OF NOTES

Our conversations delved into two issues that have bearing on any augmentations to progress notes:

Preserving the original text. While patients spoke about augmentations to their notes, we recognized they often suggested modifications that wouldn't alter the original text, such as hiding rather than deleting irrelevant sections (Section 6.1). The progress notes' value partly lies in revealing clinicians' straightforward evaluations ($N = 3$, see Section 5). P9 emphasized the significance of understanding her diagnosis and life expectancy from the clinician's viewpoint, stating, "*To be quite honest, I'm looking for facts. [Facts] are reassuring, even if [they aren't said] in a reassuring tone.*" Hence, we posit that enhancements to the reading experience must maintain access to the original text to preserve these benefits.

Perspectives on incorporating AI. Given contemporary concerns about hallucinations in AI-generated texts, we anticipated

patients would uniformly be resistant to our suggestions of using AI to support the reading of progress notes. While some patients were indeed resistant to the idea, others welcomed it ($N = 4$). For instance, P2 conveyed great enthusiasm for the automatically generated summaries of notes in the *summary* mockup, telling us, "*I'm a believer in artificial intelligence… if you can get from the left paragraph [a note's assessment and plan] to the right paragraph [generated summary] using AI, I think that would be fantastic.*" Patients appreciated the idea of AIs that explained the reasoning behind a clinician's decisions regarding patient care (P14) or framed content in the note "*to make it less alarming*" (P9).

Other patients were more skeptical, anticipating that AIs would make errors. P14 channeled a concern that AIs may generate incorrect interpretations of the note, stating that "*the concern is that we're relying too much on that, and taking something out of the mix, and it may be a different, maybe, and an inappropriate interpretation of something because it misinterpreted someone's intent.*"

Some patients therefore anticipated that AI-generated text would need to be edited (P5) or reviewed by the clinician (P14). And if clinicians did not play an active role in reviewing the text before it was shown to the patient, they should at least be made aware of the AI-generated text after the fact ($N = 2$), so that they know what information patients are using to self-inform about their care.

8 DISCUSSION

8.1 Summary of findings

Our study characterized three patient needs that arise when reading progress notes. These needs are *directing attention* to aspects of a note that are important and away from those that are irrelevant, supporting *phrase-level understanding*, and helping patients put their notes in context by helping them *trace lines of reasoning*.

We then identified eight opportunities to address these needs, by incorporating context-appropriate definitions of jargon, summarizing irrelevant or verbose sections of the note, emphasizing important information, de-emphasizing less important information, filling in the gaps with necessary interpretations, linking patients to providers with complementary perspectives, connecting patients with others experiencing similar health journeys, and incorporating comforting messages from clinicians in sections that might
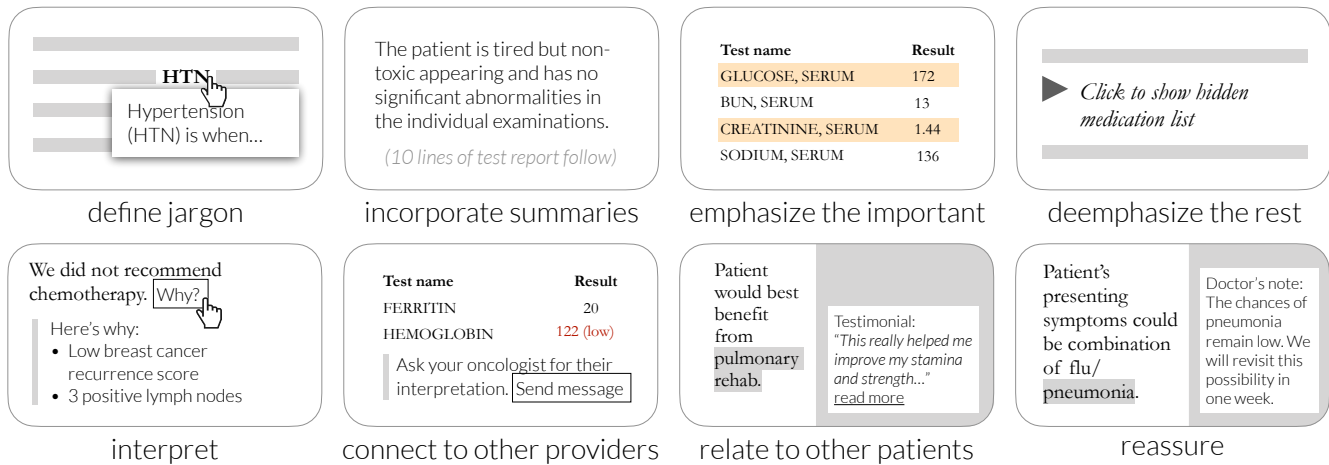
Figure 5: A visual summary of the opportunities that presented in our study for intelligent interfaces to help patients read their progress notes. Each pane corresponds to an opportunity detailed in Section 6. Some opportunities represent evolutions of ideas from the mockups we brought into the study (*incorporate summaries*, *interpret*, *relate to other patients*, and *reassure*).

cause concern. A visual summary of these opportunities appears in Figure 5, many of which show passages from patients in the study augmented with affordances that might have helped them.

Our contribution is the crystallization of the three needs, and supporting insights that deepen our knowledge beyond prior research. Organized by need, these insights are:

Directing Attention. Our study describes issues of reading as relating not just to vocabulary and interpretation, but in also directing one's attention. We characterize what patients wish to find in a note—such as lab or test results, assessments, and plans—and the passages they find to be distracting—such as medication lists, past medical history, and physical examination sections. We refine what it would mean to provide attentional guidance for notes by suggesting the application of affordances for emphasis (e.g., as explored by Fok et al. [20]) in a way that spotlights key information while preserving access to the original note content.

Phrase-Level Understanding. Our study begins by reproducing what is known—namely, that the terminology in notes can be hard for patients to understand, and that such terminology includes both words and acronyms. Then, we observe that difficulty arises not just from understanding individual terms, but in understanding terminology-dense passages—an observation made in prior research studying biomedical research papers [5], though not yet in progress notes. We characterize one challenge as the overloaded meanings of acronyms, and observe cases where patients selected the wrong meaning for an acronym. Furthermore, we bring new nuance to the notion of generating useful definitions, recognizing that different patients will need definitions for different terms, and that definitions may need to be tailored to the contexts terms are used in.

Tracing Lines of Reasoning. We deepen our understanding of how to help patients interpret their notes. Prior work stops short at suggesting patient education for interpreting health-related numbers [21] and recommending patients be able to identify parts of

notes that signal changes in treatment [17]. We identify key interpretation targets: next steps in care, reasons for labs or tests, test implications, and connections between diagnoses and results. We suggest affordances that may require going beyond traditional biomedical QA approaches (e.g., [30, 85]), advocating for a interpretation support that integrates passage analysis, clarification of concerning terms, incorporation of second opinions, and connecting patients to the experiences of others.

8.2 Envisioning the "explainable note"

Drawing on our findings, we see an opportunity to raise the profile of the progress note as the target of study within the HCI community. In this section, we introduce the notion of an *explainable note* to draw together our findings into a single vision of an enhanced patient experience. An explainable note is an augmented progress note that provides integrated support for directing attention, supporting phrase-level understanding, and tracing lines of reasoning. It does so on the basis of patient goals and knowledge. The term "explainable" is used both in the general sense of promoting understanding, and also in the sense often used in the computational sciences of elucidating the reasoning between decisions, such as health assessments. A note is made explainable through some combination of augmented reading experiences for patients and assisted note-writing experiences for clinicians.

Our vision of an explainable note is aspirational here, in that some features may be out of reach, though we hope consideration of them can help stage advances in biomedical NLP and conversations around its use in clinical settings. To convey what it might look like to have a medical system that supported explainable notes end-to-end, we reintroduce a situation from Section 6.3 where P15 encountered a phrase their clinician had highlighted in red and which, to them, subsequently seemed alarming. With an explainable notes system, P15's clinician, who we will call Dr. K, receives support from their note-writing tool. As they highlight the finding in P15's slit lamp eye exam result in red, their note editor prompts,

"Would you like to add an explanatory note for the patient?" The AI suggests the following text: "The red text indicates findings consistent with early changes in the lens of your eye, suggestive of cataract formation." This is a concerning message perhaps, though it is concrete in making clear what the risk is to the patient. Dr. K reviews the message, adding a note that the team will discuss management options during P15's next visit.

Later, P15 opens their note in an augmented reading interface. Their attention is guided to some of the most important information in the note; the past medical history is hidden away behind a clickable label. The slit lamp eye exam is not, however. P15's attention is directed, as before, to Dr. K's red highlight in the table of exam results. This time P15 notices the table cell is clickable. They click on the cell to see the definition of the phrase "nuclear sclerosis" and see that it is "the hardening of the eye's lens, often a result of aging." They click Dr. K's accompanying note, to see a deeper interpretation of the result as reason for concern, though one that the clinician is tracking nonetheless. Now P15 knows of it as well. The explainable note has helped them rapidly consume some of the most important information in their note.

8.3 Towards the explainable note

What would it take to develop an explainable note? Below, we take stock of the affordances of the explainable note, and the kinds of developments needed in HCI and AI to bring them about.

8.3.1 Cross-cutting hazards. A successful design of an explainable note needs to acknowledge several tensions:

AI inaccuracies. One of the risks of using AI to augment medical text is its potential to generate inaccurate medical information [14, 24, 51]. Should generative AI produce inaccurate patient-facing information, patients could make incorrect conclusions about their care. The risk could disproportionately disadvantage those marginalized by the health system, whom may not have the health literacy necessary to critically evaluate generations. Given the potential for harm, some have advocated that generated AI tools undergo the same kind of review as medical devices, especially when being considered for use in health applications [24]. We believe that responsible development and implementation of explainable notes will incorporate methods for minimizing inaccuracies and mitigating their harm.

Clinician burden. Today, clinicians face considerable documentation burden and high levels of burnout due to their increasing clinical care needs and the cognitive burden associated with clinical information management [82]. For explainable notes to become a sustainable fixture of current health systems, they would need to be developed in a way that minimize additional burden to the clinician. Ideally they would reduce the need for patient portal messaging and improve therapy compliance and clinic follow up rates, all of which would improve the throughput of the health care system.

Patient burden. In the past, augmentations to texts have sometimes introduced new burdens, like cognitive overhead of understanding the augmentations [16].

Private information. Patient's progress notes are considered protected data. As such, AI tools used in producing explainable notes must be incorporated into health information systems in a way that gives providers and patients complete control over their privacy and usage of their data.

8.3.2 Plotting a way forward. What trends in research suggest a way forward in developing explainable notes, and what problems need to be solved? Below, we describe a number of challenge problems, and how they might be addressed by HCI and AI research.

Definition. An explainable note should detect terms that a patient would like to understand, and provide context-sensitive definitions of those terms. Techniques for generating definitions have continued to mature in recent years. Case studies of LLMs have shown some ability to define medical terms [46]. Generated definitions have been rated as somewhat high quality in human evaluations [28], and some models have even been able to tailor the level of complexity of definitions [4]. Recent HCI research suggests affordances for definitions, i.e., with overlay tooltips [5, 26]; what should be better understood is how to find what terms a patient would want to know, and tailor definitions to the patient and context.

Emphasis. For an explainable note to direct a reader's attention, it needs to selectively emphasize and deemphasize content. Emphasis and deemphasis can be accomplished, in part, with AI techniques. Fok et al. [20], for instance, developed models for detecting important passages in scientific texts. Techniques for aspect-based text summarization (e.g., [83]) could be combined with chain-of-thought generation approaches [79] to extract significant passages from a note. As Fok et al. [20] note, such techniques may need to achieve adequate coverage of a note to appear reliable. Moreover, it may be that individual patients would desire control over which parts of their notes are highlighted as their care evolves.

Longer generations. Our proposed explainable note incorporates longer-form AI-generated content, like summaries of dense passages, interpretations of lab results, and (clinician-edited) messages of reassurance. Methods for long-form text generation have advanced considerably in recent years. For instance, ChatGPT has been shown to be capable of generating high-quality, highly-accurate plain language summaries of radiology reports [31, 47], and one recent model for biomedical QA [69] has produced answers that are judged by clinicians to be superior to clinicians' answers.

The challenges of incorporating longer generations into a note are two-fold. First, they are likely to be context-dependent, so they would need to be generated and verified for each patient. Second, patients are likely to need many of them—i.e., perhaps one per paragraph, and one per test result. What can HCI and AI do to maximize readability, minimize inaccuracies, and mitigate risks of inaccuracies? We describe several considerations below.

Factuality checking. Inaccuracies should be removed to the extent possible by automated means. This could be done with techniques for reducing hallucinations, e.g., by detecting conflicting information in generated texts [39, 84], employing retrieval-augmented generation [44], or filtering out inaccurate texts by using models to classify factuality [49].

Organizational and social mitigations. Explainable notes should be part of health systems that promote their responsible use. Access to an explainable note might be contingent. It could be initially

limited to low-stakes medical visits, and gated until a patient is briefed by a clinician and they have demonstrated care in adhering to the recommended plan of care.

Clinician in the loop. Perhaps the most challenging question with explainable notes is what is asked of clinicians. Validation from medical professionals is necessary for AI-generated medical information when inaccuracies pose threats to patients. Future research should more clearly characterize the kinds of generations, and the circumstances, in which generated medical information would pose such a threat. For instance, selective emphasis of text may have less potential to mislead than a retrieved definition, and yet less than a generated definition, or a generated paragraph.

When threats are posed and patients are not prepared for them, a clinician would need to help create the note. In these cases, what could reduce clinician burden? Text—including the original note—could be produced with a collaborative text editing model (e.g., [58, 66]). The writing interface could be extended with affordances to mark passages of concern and afford rapid verification (e.g., [40]). These affordances might add clinician burden, though these burdens can be reduced. Clinicians would see better return should some of the assistance they provide be made applicable across notes, e.g., if they were given tools to author deterministic rules for generating interpretations of test results that had consistent interpretations, or definitions that might apply well to patients within a particular area of care and a specific level of health literacy.

8.4 Limitations

Several aspects of our study limit the generality of the conclusions that we make. A first limitation is that our sample skews towards white, retired older adults enrolled in a specific urban health system. Our results likely underrepresent experiences of younger patients, non-white patients, rural patients, those enrolled in other health systems, and those with limited access to health care. Another bias arises from participants' membership in a patient advisory council; these patients may have a disproportionately vested interest in reading their progress notes and a higher level of health literacy than other patients. Additionally, as members of an advisory council for a health system, members might be particularly keen on changing the health system or open to technology advances therein, where other patients might be more resistant to change. Our characterization of tensions and opportunities to improve note reading should be considered as representative of the patients interviewed, though not necessarily comprehensive.

A second limitation is that patients' ideas for future note-reading interfaces were potentially influenced by the mockups we showed. Exposure to these mockups might have incurred an anchoring effect, leading patients to focus more on the solutions we presented rather than exploring other, perhaps more beneficial, possibilities. Additionally, patients only described how they would use interfaces hypothetically, rather than reflecting on actual use. To mitigate anchoring biases, we encouraged patients to review their notes prior to engaging with the mockups, aiming for discussions that were informed by their actual, recent experiences with note-reading. We also note that patients volunteered ideas that did not appear in our mockups, such as tooltips that revealed definitions of terms, and were candid about the mockups' limitations.

A third limitation is that we have not examined the perspectives of clinicians and other stakeholders in arriving at our recommendations. Our patient-centric view of challenges and potential solutions is intentional, but also leaves for future work any examination of organizational realities that will necessarily influence any changes to interfaces for reading or writing notes.

8.5 Positionality statement

As always in research, our own identities and backgrounds as researchers shape the questions we asked and conclusions that we drew. As a research team, we are all U.S. citizens, representing a limited segment of the population of patients across healthcare systems globally. While we do represent a diverse group of races and ethnicities, we have stable socioeconomic status, high levels of digital and health literacy, and high levels of healthcare access. Furthermore, our recommendations are focused on envisioning improvements to medical interfaces. While our author team does include a medical professional, this focus is motivated by the lead authors' primary expertise in systems-based HCI research. We acknowledge that the challenges we address in clinician-patient communication deserve consideration through other avenues, such as patient education, clinician training, and changes to the healthcare system at large. These alternative solutions may be preferable, should they eliminate the need for new systems altogether—for example, perhaps process adjustments could lead to notes being written in a way that is more patient-centric in the first place. The recommendations of this paper — and their limitations — should therefore be considered as our specific team's system-centric outlook on how issues in clinician-patient communication could be resolved.

8.6 Future work

We foresee several exciting directions for future research that builds on the conclusions of our study.

Expanding the community of patients under study. Similar studies with a broader population would expand the insights from our study. Of particular importance is to understand the reading needs of those who are marginalized by the health system, have less education, or are unemployed [60]. The perspectives of patients in these groups may reveal additional opportunities and constraints for augmenting notes in a way that benefits readers.

Incorporating the clinician's perspective. Our recommendations span across reading interfaces and writing interfaces. The acceptability of writing interfaces will depend on fit for the clinicians who write notes. Future studies centered around clinicians are crucial for designing interfaces that help patients read progress notes without them being too much of a burden to write.

Contextual design studies. Patients will likely have deeper, better-grounded ideas for how to improve the note reading experience if they are given the ability to use prototype designs on their own notes, at the times they would like to read them. Implementing these interfaces on a smaller scale is a crucial step for understanding patient behaviors with greater granularity.

Designing augmented note reading and writing interfaces. Figure 5 outlines a set of affordances of an explainable note, and

Section 8.3.2 describes challenge problems in HCI and AI needed to bring them about. Future research should crystallize the affordances and take on the challenge problems.

9 CONCLUSION

By identifying the barriers patients encounter in reading progress notes and exploring opportunities for improvements, our study lays the groundwork for systems of intelligent interfaces that enhance patients' comprehension of their progress notes. Implementing these proposed changes can lead to more informed and empowered patients, fostering a positive impact on patient-clinician relationships and overall healthcare outcomes.

ACKNOWLEDGMENTS

REFERENCES

[1] Houtan Aghili, Richard A. Mushlin, Rose M. Williams, and Jeffrey S. Rose. 1997. Progress notes model. In *Proceedings of the American Medical Informatics Association Annual Fall Symposium*. American Medical Informatics Association, 12–16.

[2] Brian G. Arndt, John W. Beasley, Michelle D. Watkinson, Jonathan L. Temte, Wen-Jan Tuan, Christine A. Sinsky, and Valerie J. Gilchrist. 2017. Tethered to the EHR: Primary Care Physician Workload Assessment Using EHR Event Log Data and Time-Motion Observations. *The Annals of Family Medicine* 15, 5 (2017), 419–426.

[3] Lord Atkin. 1932. Donoghue v. Stevenson. *UKHL* 100 (1932), 26.

[4] Tal August, Katharina Reinecke, and Noah A. Smith. 2022. Generating Scientific Definitions with Controllable Complexity. In *Proceedings of Annual Meeting of the Association for Computational Linguistics*. ACL, 8298–8317.

[5] Tal August, Lucy Lu Wang, Jonathan Bragg, Marti A. Hearst, Andrew Head, and Kyle Lo. 2022. Paper Plain: Making medical research papers approachable to healthcare consumers with natural language processing. *ACM Transactions on Computer-Human Interaction* (2022).

[6] Molly Baldry, Carol Cheal, Brian Fisher, Myra Gillett, and Val Huet. 1986. Giving patients their own records in general practice: experience of patients and staff. *British Medical Journal (Clinical Research Ed)* 292, 6520 (1986), 596–598.

[7] John W. Beasley, Tosha B. Wetterneck, Jon Temte, Jamie A. Lapin, Paul Smith, A. Joy Rivera-Rodriguez, and Ben-Tzion Karsh. 2011. Information chaos in primary care: implications for physician performance and patient safety. *The Journal of the American Board of Family Medicine* 24, 6 (2011), 745–751.

[8] Sigall K. Bell, Tom Delbanco, Joann G. Elmore, Patricia S. Fitzgerald, Alan Fossa, Kendall Harcourt, Suzanne G. Leveille, Thomas H. Payne, Rebecca A. Stametz, Jan Walker, et al. 2020. Frequency and Types of Patient-Reported Errors in Electronic Health Record Ambulatory Care Notes. *The Journal of the American Medical Association Network Open* 3, 6 (2020), e205867.

[9] Sigall K. Bell, Macda Gerard, Alan Fossa, Tom Delbanco, Patricia H. Folcarelli, Kenneth E. Sands, Barbara Sarnoff Lee, and Jan Walker. 2017. A patient feedback reporting tool for OpenNotes: implications for patient-clinician safety and quality partnerships. *British publisher of medical journals (BMJ) quality & safety* 26, 4 (2017), 312–322.

[10] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. *Qualitative HCI Research: Going Behind the Scenes*. Morgan & Claypool Publishers.

[11] Carrie J. Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S. Corrado, Martin C. Stumpe, and Michael Terry. 2019. Human-Centered Tools for Coping with Imperfect Algorithms During Medical Decision-Making. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 4.

[12] Hannah Chimowitz, Macda Gerard, Alan Fossa, Fabienne Bourgeois, and Sigall K. Bell. 2018. Empowering informal caregivers with health information: OpenNotes as a safety strategy. *The Joint Commission Journal on Quality and Patient Safety* 44, 3 (2018), 130–136.

[13] Soon Hau Chua, Toni-Jan Keith Palma Monserrat, Dongwook Yoon, Juho Kim, and Shengdong Zhao. 2017. Korero: Facilitating complex referencing of visual materials in asynchronous discussion interface. *Proceedings of the ACM on Human-Computer Interaction* 1 (2017), 34.

[14] Jan Clusmann, Fiona R. Kolbinger, Hannah Sophie Muti, Zunamys I. Carrero, Jan-Niklas Eckardt, Narmin Ghaffari Laleh, Chiara Maria Lavinia Löffler, Sophie-Caroline Schwarzkopf, Michaela Unger, Gregory P. Veldhuizen, Sophia J. Wagner, and Jakob Nikolas Kather. 2023. The future landscape of large language models in medicine. *Communications Medicine* 3, 1 (2023), 141.

[15] Tom Delbanco, Jan Walker, Jonathan D. Darer, Joann G. Elmore, Henry J. Feldman, Suzanne G. Leveille, James D. Ralston, Stephen E. Ross, Elisabeth Vodicka, and Valerie D. Weber. 2010. Open Notes: Doctors and Patients Signing On. *Annals of Internal Medicine* 153, 2 (2010), 121–125.

[16] Diana DeStefano and Jo-Anne LeFevre. 2007. Cognitive load in hypertext reading: A review. *Computers in human behavior* 23, 3 (2007), 1616–1641.

[17] Mark A. Earnest, Stephen E. Ross, Loretta Wittevrongel, Laurie A. Moore, and Chen-Tan Lin. 2004. Use of a patient-accessible electronic medical record in a practice for congestive heart failure: patient and physician experiences. *Journal of the American Medical Informatics Association* 11, 5 (2004), 410–417.

[18] Tobias Esch, Roanne Mejilla, Melissa Anselmo, Beatrice Podtschaske, Tom Delbanco, and Jan Walker. 2016. Engaging patients through open notes: an evaluation using mixed methods. *British Medical Journal Open* 6, 1 (2016), e010034.

[19] Sarah Faisal, Ann Blandford, and Henry W. W. Potts. 2013. Making sense of personal health information: challenges for information visualization. *Health informatics journal* 19, 3 (2013), 198–217.

[20] Raymond Fok, Hita Kambhamettu, Luca Soldaini, Jonathan Bragg, Kyle Lo, Marti A. Hearst, Andrew Head, and Daniel S. Weld. 2023. Scim: Intelligent Skimming Support for Scientific Papers. In *Proceedings of the International Conference on Intelligent User Interfaces*. ACM, 476–490.

[21] Perry M. Gee, Debora A. Paterniti, Deborah Ward, and Lisa M. Soederberg Miller. 2015. e-Patients perceptions of using personal health records for self-management support of chronic illness. *CIN: Computers, Informatics, Nursing* 33, 6 (2015), 229–237.

[22] Macda Gerard, Alan Fossa, Patricia H. Folcarelli, Jan Walker, and Sigall K. Bell. 2017. What Patients Value About Reading Visit Notes: A Qualitative Inquiry of Patient Experiences With Their Health Information. *Journal of Medical Internet Research* 19, 7 (2017), e237.

[23] Traber Davis Giardina, Helen Haskell, Shailaja Menon, Julia Hallisy, Frederick S. Southwick, Urmimala Sarkar, Kathryn E. Royse, and Hardeep Singh. 2018. Learning From Patients' Experiences Related to Diagnostic Errors is Essential for Progress in Patient Safety. *Health Affairs* 37, 11 (2018), 1821–1827.

[24] Stephen Gilbert, Hugh Harvey, Tom Melvin, Erik Vollebregt, and Paul Wicks. 2023. Large language model AI chatbots require approval as medical devices. *Nature Medicine* 29 (2023), 2396–2398.

[25] Lisa V. Grossman, Ruth Masterson Creber, Susan Restaino, and David K. Vawdrey. 2017. Sharing Clinical Notes with Hospitalized Patients via an Acute Care Portal. In *American Medical Informatics Association Annual Symposium Proceedings*, Vol. 2017. American Medical Informatics Association, 800–809.

[26] Andrew Head, Kyle Lo, Dongyeop Kang, Raymond Fok, Sam Skjonsberg, Daniel S. Weld, and Marti A. Hearst. 2021. Augmenting Scientific Papers with Just-in-Time, Position-Sensitive Definitions of Terms and Symbols. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–18.

[27] Hendrik Heuer and Elena L. Glassman. 2023. Accessible Text Tools: Where They Are Needed & What They Should Look Like. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–7.

[28] Jie Huang, Hanyin Shao, Kevin Chen-Chuan Chang, Jinjun Xiong, and Wen-mei Hwu. 2022. Understanding Jargon: Combining Extraction and Generation for Definition Modeling. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*. ACL, 3994–4004.

[29] Sean S. Huang, Shane P. Stenner, and S. Trent Rosenbloom. 2023. The 21st Century Cures Act Information Blocking Rule in Post-Acute Long-Term Care. *Journal of the American Medical Directors Association* 25 (2023), 58–60.

[30] Farnaz Jahanbakhsh, Elnaz Nouri, Robert Sim, Ryen W. White, and Adam Fourney. 2022. Understanding Questions that Arise When Working with Business Documents. *Proceedings of the ACM on Human-Computer Interaction* 6 (2022).

[31] Katharina Jeblick, Balthasar Schachtner, Jakob Dexl, Andreas Mittermeier, Anna Theresa Stüber, Johanna Topalis, Tobias Weber, Philipp Wesp, Bastian Oliver Sabel, Jens Ricke, et al. 2023. ChatGPT Makes Medicine Easy to Swallow: An Exploratory Case Study on Simplified Radiology Reports. *European Radiology* (2023), 1–9.

[32] Zhuoren Jiang, Liangcai Gao, Ke Yuan, Zheng Gao, Zhi Tang, and Xiaozhong Liu. 2018. Mathematics Content Understanding for Cyberlearning via Formula Evolution Map. In *Proceedings of the ACM International Conference on Information and Knowledge Management*. ACM, 37–46.

[33] Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific Data* 3, 1 (2016).

[34] Hyeonsu B. Kang, Joseph Chee Chang, Yongsung Kim, and Aniket Kittur. 2022. Threddy: An Interactive System for Personalized Thread-based Exploration and

Organization of Scientific Literature. In *Proceedings of the Symposium on User Interface Software and Technology*. ACM, 1–15.

[35] Hyeonsu B. Kang, Tongshuang Wu, Joseph Chee Chang, and Aniket Kittur. 2023. Synergi: A Mixed-Initiative System for Scholarly Synthesis and Sensemaking. In *Proceedings of the Symposium on User Interface Software and Technology*. 1–19.

[36] Amro Khasawneh, Ian Kratzke, Karthik Adapa, Lawrence Marks, and Lukasz Mazur. 2022. Effect of Notes' Access and Complexity on OpenNotes' Utility. *Applied Clinical Informatics* 13, 05 (2022), 1015–1023.

[37] Tae Soo Kim, Matt Latzke, Jonathan Bragg, Amy X. Zhang, and Joseph Chee Chang. 2023. Papeos: Augmenting Research Papers with Talk Videos. In *Proceedings of the Symposium on User Interface Software and Technology*. 1–19.

[38] Ross Koppel. 2022. Healthcare Information Technology's Relativity Challenges: Distortions Created by Patients' Physical Reality versus Clinicians' Mental Models and Healthcare Electronic Records. *Qualitative Sociology Review* 18, 4 (2022), 92–108.

[39] Philippe Laban, Tobias Schnabel, Paul N. Bennett, and Marti A. Hearst. 2022. SummaC: Re-visiting NLI-based models for inconsistency detection in summarization. *Transactions of the Association for Computational Linguistics* 10 (2022), 163–177.

[40] Philippe Laban, Jesse Vig, Marti A. Hearst, Caiming Xiong, and Chien-Sheng Wu. 2023. Beyond the Chat: Executable and Verifiable Text-Editing with LLMs. (2023). arXiv:2309.15337 [cs.CL]

[41] Barbara D. Lam, Fabienne Bourgeois, Zhiyong J. Dong, and Sigall K. Bell. 2021. Speaking up about patient-perceived serious visit note errors: Patient and family experiences and recommendations. *Journal of the American Medical Informatics Association* 28, 4 (2021), 685–694.

[42] Byungjoo Lee, Olli Savisaari, and Antti Oulasvirta. 2016. Spotlights: Attention-Optimized Highlights for Skim Reading. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 5203–5214.

[43] Laura D. Leonard, Ben Himelhoch, Victoria Huynh, Dulcy Wolverton, Kshama Jaiswal, Gretchen Ahrendt, Sharon Sams, Ethan Cumbler, Richard Schulick, and Sarah E. Tevis. 2022. Patient and clinician perceptions of the immediate release of electronic health information. *The American Journal of Surgery* 224, 1 (2022), 27–34.

[44] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems* 33 (2020), 9459–9474.

[45] Claudio D.G. Linhares, Daniel M. Lima, Jean R. Ponciano, Mauro M. Olivatto, Marco A. Gutierrez, Jorge Poco, Caetano Traina, and Agma Juci Machado Traina. 2022. ClinicalPath: A Visualization Tool to Improve the Evaluation of Electronic Health Records in Clinical Decision-Making. *IEEE Transactions on Visualization and Computer Graphics* 29 (2022), 4031–4046.

[46] Renqian Luo, Liai Sun, Yingce Xia, Tao Qin, Sheng Zhang, Hoifung Poon, and Tie-Yan Liu. 2022. BioGPT:Generative Pre-trained Transformer for Biomedical Text Generation and Mining. *Briefings in Bioinformatics* 23, 6 (2022), bbac409.

[47] Qing Lyu, Josh Tan, Michael E. Zapadka, Janardhana Ponnatapura, Chuang Niu, Kyle J. Myers, Ge Wang, and Christopher T. Whitlow. 2023. Translating radiology reports into plain language using ChatGPT and GPT-4 with prompt learning: results, limitations, and potential. *Visual Computing for Industry, Biomedicine, and Art* 6, 1 (2023), 9.

[48] Lena Mamykina, Elizabeth Mynatt, Patricia Davidson, and Daniel Greenblatt. 2008. MAHI: investigation of social scaffolding for reflective thinking in diabetes management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 477–486.

[49] Potsawee Manakul, Adian Liusie, and Mark Gales. 2023. SelfCheckGPT: Zero-Resource Black-Box Hallucination Detection for Generative Large Language Models. In *The Conference on Empirical Methods in Natural Language Processing*. ACL, 9004–9017.

[50] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proceedings of the ACM on Human-Computer Interaction* 3 (2019).

[51] Bertalan Meskó and Eric J. Topol. 2023. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *Nature Partner Journal Digital Medicine* 6, 1 (2023), 120.

[52] Vimal K. Mishra, Robert E. Hoyt, Susan E. Wolver, Ann Yoshihashi, and Colin Banas. 2019. Qualitative and Quantitative Analysis of Patients' Perceptions of the Patient Portal Experience with OpenNotes. *Applied Clinical Informatics* 10, 01 (2019), 10–18.

[53] Luke Murray, Divya Gopinath, Monica Agrawal, Steven Horng, David Sontag, and David R. Karger. 2021. MedKnowts: Unified Documentation and Information Retrieval for Electronic Health Records. In *Proceedings of the Symposium on User Interface Software and Technology*. ACM, 1169–1183.

[54] Avinash Murugan, Holly Gooding, Jordan Greenbaum, Jeanne Boudreaux, Reena Blanco, Arin Swerlick, Cary Sauer, Steven Liu, Amina Bhatia, Alexis Carter, Meredith M. Burris, Lauren Becker, Lashandra Abney, Sharon O'Brien, Shane Webb, Melissa Popkin, Herb Williams, Desiree Jennings, and Evan W. Orenstein.

2022. Lessons Learned from OpenNotes Learning Mode and Subsequent Implementation across a Pediatric Health System. *Applied Clinical Informatics* 13 (2022), 113–122.

[55] Sheshera Mysore, Mahmood Jasim, Haoru Song, Sarah Akbar, Andre Kenneth Chase Randall, and Narges Mahyar. 2023. How Data Scientists Review the Scholarly Literature. In *Proceedings of the Conference on Human Information Interaction and Retrieval*. ACM, 137–152.

[56] Jenny Park, Somnath Saha, Brant Chee, Janiece Taylor, and Mary Catherine Beach. 2021. Physician Use of Stigmatizing Language in Patient Medical Records. *The Journal of the American Medical Association Network Open* 4, 7 (2021).

[57] Napol Rachatasumrit, Jonathan Bragg, Amy X. Zhang, and Daniel S. Weld. 2022. CiteRead: Integrating Localized Citation Contexts into Scientific Paper Reading. In *International Conference on Intelligent User Interfaces*. ACM, 707–719.

[58] Vipul Raheja, Dhruv Kumar, Ryan Koo, and Dongyeop Kang. 2023. CoEdIT: Text Editing by Task-Specific Instruction Tuning. In *Findings of the Association for Computational Linguistics: EMNLP*. ACL, 5274–5291.

[59] Tera L. Reynolds, Nida Ali, Emma McGregor, Trish O'Brien, Christopher Longhurst, Andrew L. Rosenberg, Scott E. Rudkin, and Kai Zheng. 2017. Understanding Patient Questions about their Medical Records in an Online Health Forum: Opportunity for Patient Portal Design. In *American Medical Informatics Association Annual Symposium Proceedings*, Vol. 2017. American Medical Informatics Association, 1468.

[60] Joseph Root, Natalia V. Oster, Sara L. Jackson, Roanne Mejilla, Jan Walker, and Joann G. Elmore. 2016. Characteristics of Patients Who Report Confusion After Reading Their Primary Care Clinic Notes Online. *Health Communication* 31, 6 (2016), 778–781.

[61] Stephen E. Ross and Chen-Tan Lin. 2003. The Effects of Promoting Patient Access to Medical Records: A Review. *Journal of the American Medical Informatics Association* 10, 2 (2003), 129–138.

[62] Adam Rule, Steven Bedrick, Michael F. Chiang, and Michelle R. Hribar. 2021. Length and Redundancy of Outpatient Progress Notes Across a Decade at an Academic Medical Center. *The Journal of the American Medical Association Network Open* 4, 7 (2021).

[63] Adam Rule, Isaac H. Goldstein, Michael F. Chiang, and Michelle R. Hribar. 2020. Clinical Documentation as End-User Programming. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Association of Computing Machinery, 1–13.

[64] Rohit B. Sangal, Emily Powers, Craig Rothenberg, Chima Ndumele, Andrew Ulrich, Allen Hsiao, and Arjun K. Venkatesh. 2021. Disparities in Accessing and Reading Open Notes in the Emergency Department Upon Implementation of the 21st Century CURES Act. *Annals of Emergency Medicine* 78, 5 (2021), 593–598.

[65] Urmimala Sarkar, Andrew J. Karter, Jennifer Y. Liu, Nancy E. Adler, Robert Nguyen, Andrea Lopez, and Dean Schillinger. 2010. The Literacy Divide: Health Literacy and the Use of an Internet-Based Patient Portal in an Integrated Health System—Results from the Diabetes Study of Northern California (DISTANCE). *Journal of Health Communication* 15, S2 (2010), 183–196.

[66] Timo Schick, Jane A. Yu, Zhengbao Jiang, Fabio Petroni, Patrick Lewis, Gautier Izacard, Qingfei You, Christoforos Nalmpantis, Edouard Grave, and Sebastian Riedel. 2023. PEER: A Collaborative Language Model. In *The International Conference on Learning Representations*.

[67] Jeffrey L. Schnipper, Jeffrey A. Linder, Matvey B. Palchuk, Jonathan S. Einbinder, Qi Li, Anatoly Postilnik, and Blackford Middleton. 2008. "Smart Forms" in an Electronic Medical Record: documentation-based clinical decision support to improve disease management. *Journal of the American Medical Informatics Association* 15, 4 (2008), 513–523.

[68] Jessica Schroeder, Jane Hoffswell, Chia-Fang Chung, James Fogarty, Sean Munson, and Jasmine Zia. 2017. Supporting Patient-Provider Collaboration to Identify Individual Triggers using Food and Symptom Journals. In *Proceedings of the Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 1726–1739.

[69] Karan Singhal, Tao Tu, Juraj Gottweis, Rory Sayres, Ellery Wulczyn, Le Hou, Kevin Clark, Stephen Pfohl, Heather Cole-Lewis, Darlene Neal, et al. 2023. Towards Expert-Level Medical Question Answering with Large Language Models. *arXiv preprint arXiv:2305.09617* (2023).

[70] David Spiro and Fred Heidrich. 1983. Lay understanding of medical terminology. *Journal of Family Practice* 17, 2 (1983), 277–9.

[71] Nicole Sultanum, Michael Brudno, Daniel Wigdor, and Fanny Chevalier. 2018. More Text Please! Understanding and Supporting the Use of Visualization for Clinical Text Overview. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, 1–13.

[72] Nicole Sultanum, Farooq Naeem, Michael Brudno, and Fanny Chevalier. 2022. ChartWalk: Navigating large collections of text notes in electronic health records for clinical chart review. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2022), 1244–1254.

[73] Nicole Sultanum, Devin Singh, Michael Brudno, and Fanny Chevalier. 2018. Doccurate: A Curation-Based Approach for Clinical Text Visualization. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2018), 142–151.

[74] Craig S. Tashman and W. Keith Edwards. 2011. LiquidText: A Flexible, Multitouch Environment to Support Active Reading. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3285–3294.

[75] Konrad Tollmar, Frank Bentley, and Cristobal Viedma. 2012. Mobile Health Mashups: Making sense of multiple streams of wellbeing and contextual data for presentation on a mobile device. In *International Conference on Pervasive Computing Technologies for Healthcare*. IEEE, 65–72.

[76] Peter S Uzelac and Richard W Moon. 2005. *SOAP for Internal Medicine*. Lippincott Williams & Wilkins.

[77] Jan Walker, Michael Meltsner, and Tom Delbanco. 2015. US experience with doctors and patients sharing clinical notes. *British Medical Journal* 350 (2015).

[78] Jixuan Wang, Jingbo Yang, Haochi Zhang, Helen Lu, Marta Skreta, Mia Husić, Aryan Arbabi, Nicole Sultanum, and Michael Brudno. 2022. PhenoPad: Building AI enabled note-taking interfaces for patient encounters. *Nature Partner Journal Digital Medicine* 5, 1 (2022), 12.

[79] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V. Le, Denny Zhou, et al. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *Advances in Neural Information Processing Systems* 35 (2022), 24824–24837.

[80] Lauren Wilcox, Jie Lu, Jennifer Lai, Steven Feiner, and Desmond Jordan. 2010. Physician-Driven Management of Patient Progress Notes in an Intensive Care Unit. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1879–1888.

[81] Jesse O. Wrenn, Daniel M. Stein, Suzanne Bakken, and Peter D. Stetson. 2010. Quantifying clinical narrative redundancy in an electronic health record. *Journal of the American Medical Informatics Association* 17, 1 (2010), 49–53.

[82] Qi Yan, Zheng Jiang, Zachary Harbin, Preston H. Tolbert, and Mark G. Davies. 2021. Exploring the relationship between electronic health records and provider burnout: a systematic review. *Journal of the American Medical Informatics Association* 28, 5 (2021), 1009–1021.

[83] Xianjun Yang, Yan Li, Xinlu Zhang, Haifeng Chen, and Wei Cheng. 2023. Exploring the Limits of ChatGPT for Query or Aspect-based Text Summarization. (2023).

[84] Xiang Yue, Boshi Wang, Ziru Chen, Kai Zhang, Yu Su, and Huan Sun. 2023. Automatic Evaluation of Attribution by Large Language Models. In *The Conference on Empirical Methods in Natural Language Processing*. ACL, 4615–4635.

[85] Tiancheng Zhao and Kyusong Lee. 2020. Talk to Papers: Bringing Neural Question Answering to Academic Search. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. ACL, 30–36.

A APPENDICES

A.1 Participant backgrounds

ID	Age	Gender	Ethnicity	Occupation	Reads notes	Specialty of care	Chronic
1	58	female	Asian	Housewife	never	(no condition)	
2	62	male	White	Recently retired	daily	Pulmonology and Cardiology	✓
3	84	female	White	Retired	< monthly	Ears, Nose, Throat (ENT)-related	
4	61	female	White, Some Other	Teacher and consultant	never	Ophthalmology	
5	70	female	Black/Afr. Am.	Retired	< monthly	Dermatology	
6	36	female	White	Auditor	weekly	Lymphology	✓
7	23	male	White, Asian	Clinical research coordinator	never	(no condition)	
8	78	male	Black/Afr. Am.	Retired	daily	Urology	✓
9	55	female	White	Self-employed	< monthly	Gynecology	✓
10	62	male	White	Semi-retired consultant	(no response)	Gastroenterology	✓
11	79	female	White	Retired	< monthly	Lymphology	✓
12	81	male	White	Retired attorney	< monthly	Cardiology	
13	71	male	White	Semi-retired healthcare business consultant and investment banker	< monthly	Cardiology	✓
14	78	female	White	Retired counselor	< monthly	Ophthalmology	✓

Table 1: Participant Backgrounds and Characteristics.

A.2   Mockups

Below, we show the five mockups patients were shown during the study, as described in Section 3.2. The mockups were static, still images. In the captions we describe aspects of interactivity that were envisioned for each mockup.

```
ALLERGIES:
Patient has answered NKA

FAMILY HISTORY:
Non-contributory

REVIEW OF SYSTEMS:
10-point ROS negative except as otherwise noted.

PHYSICAL EXAM:

 The patient is tired but non-toxic appearing and has no significant
abnormalities in the head, neck, cardiovascular, respiratory, abdominal,
and skin examinations.

There are no gross deformities or focal/lateralizing neurological
deficits present, and the patient is alert and oriented with a normal
mood and affect.

Click to read the full report
```

Figure 6: Summary. An augmented physical exam section of a progress note. The section is augmented with an AI-generated plain language summary of its contents. Only the generated summary is showing; the original text of the jargon-heavy section is hidden. The patient can access the original text by clicking the underlined link.

```
Assessment:
 Differential diagnoses remain broad with exact etiology still unclear.
Most concerning at this time remains pericarditis. Acute coronary
syndrome remains in differential but is less likely given the normal
cardiac enzymes and the nature of the EKG changes. PE is very unlikely
given the normal oxygenation. Dissection might be a possibility because
the patient has a family history of HTN and symmetric blood pressures.


Plan:
Serial EKG
Stat ECHO
Follow cardiac enzymes
Morphine for pain controlContinue ASA
Hold Beta Blocker as SBP less than 100 and HR 60
Consider cardiology consult depending on above workup
1 hr SPENT DISCUSSING MEDICAL CONDITIONS;
> 50% OF TIME SPENT IN COUNSELLING AND ANSWERING QUESTIONS.
```

When coming to this assessment, your doctor likely thought through the following before diagnosing you with pericarditis:

Your doctor noted that your sharp and constant chest pain could be associated with pericarditis, which is inflammation of the pericardium—the membrane enclosing the heart.

Your doctor might have taken into account your past history of hyperlipidemia and type II diabetes which increases the risk of cardiovascular diseases including pericarditis.

Figure 7: Diagnosis explanation. An augmented assessment section of a progress note. When a patient clicks on the assessment (highlighted in light blue), a side note appears describing the likely rationale behind the assessment.

```
======================================================================
Report Released Date/Time: Nov 14, 2022@09:58
Provider: KIM,TAE MIN
Specimen: SERUM. CH 1114 115
Specimen Collection Date: Nov 14, 2022@08:50
```

Test name	Result	units	Ref. range	Site Code
GLUCOSE, SERUM	188	mg/dL	74 – 118	[460]
BUN, SERUM	11	mg/dL	8 – 20	[460]
CREATININE, SERUM	1.41	mg/dL	0.70 – 1.20	[460]
SODIUM, SERUM	138	mmol/L	136 – 145	[460]
POTASSIUM, SERUM	3.4 L	mmol/L	3.6 – 5.1	[460]
CHLORIDE, SERUM	105	mmol/L	101 – 111	[460]
CARBON DIOXIDE, SERUM	22	mmol/L	22 – 31	[460]
CALCIUM, SERUM	8.1 L	mg/dL	8.4 – 10.2	[460]
TOTAL PROTEIN, SERUM	6.8	g/dL	6.0 – 8.0	[460]
ALBUMIN, SERUM	2.9 L	g/dL	3.5 – 5.0	[460]
ALKALINE PHOSPHATASE, SERUM	52	U/L	40 – 150	[460]
AST, SERUM	32	U/L	10 – 41	[460]
ALT, SERUM	23	U/L	5 – 55	[460]
MAGNESIUM, SERUM	1.5 L	mg/dL	1.6 – 2.6	[460]
CREATININE eGFR	55 L	mL/min	Ref: >=60	[460]

Your doctor ordered this panel in response to your complaint of frequent nighttime urination during your visit on 4/8/23.

Given your past medical history of diabetes, the cause for your frequent urination seems to be most correlated to the high levels of **glucose** and **creatinine**.

Looking at your lab results, these levels are above that of the normal range. This shift in glucose levels started around two weeks ago, according to your blood glucose monitor.

Your doctor will take a look at these lab results and discuss further steps with you during your visit next week.

```
Eval: eGFR results >60 are imprecise. Many variables affect the
calculated

Eval: result. Interpretation of eGFR results >60 must be monitored over
time.
```

Figure 8: Lab interpretation. An augmented lab results section of a progress note. As is common in notes, the lab results appear in a table. Rows of the table are highlighted to emphasize which labs are related to the patient's diagnosis. A generated side note offers reasoning for why the lab panel was ordered, and an interpretation of the labs.

```
Brief Hospital Course:
Ms. Kelley was taken to the operating room by Dr. Hill on 1/26/2022 for
her paraesophageal hernia with laparoscopic Nissen. She recovered in
usual fashion. A barium swallow study was done on 1/27/2022 which did not
show any leak. She had some coughing on 1/27/2022, at which time she was
resumed on home meds, and given aggressive pulmonary toilet. Pulmonology
evaluated her and did not feel she warranted bronchoscopy at that time.
The patient remained on her home oxygen. She was evaluated by PT/OT on
1/29/2022 who determined she would best benefit from pulmonary rehab. Her
cough worsened and chest xray revealed CHF. She was diuresed well,
however on 1/30/2022 developed 102 fever, was pancultured and started on
vancomycin and zosyn. She required transfer to the ICU for sepsis on
1/31/2022. She required low dose neosynephrine. She was found to have
MRSA pneumonia which resolved on IV vancomycin. Her last vancomycin
trough level was 18 on 2/2/2022. ID consulted and recommended PICC line
with IV vancomycin to continue until 2/14/2022 with CBC, Chem panel and
vanco trough 2/6/2022. The patient was transfered to the floor on
2/4/2022. She has been medically stable without fevers or hypotension on
the floor and is stable for pulmonary rehab. It is noted we do not have a
recent echo documenting LV function, and the patient did not come in with
beta blockers or ace inhibitors. She should have close outpatient follow
up with her primary care physician regarding initiation of these meds if
tolerated.
```

Here are some testimonials from patients who've experienced lung inflammation or sepsis similar to yours:

- " *50F here recovering from sepsis after a recent ICU visit ... I still feel like I have difficulty breathing sometimes and I get winded more easily than I used to. It's been really frustrating, especially when I try to exercise.*" click to expand.

- " *... I had nightmares about being back in the hospital and would wake up in a cold sweat. The fear of a relapse into respiratory distress was always in the back of my mind..*" click to expand.

Figure 9: Testimonial. In this mockup, a note is augmented to help a patient relate their experiences to those of others with the same condition. The patient can select passages describing aspects of their condition (here highlighted in light blue) to pull up a side note. The side note contains testimonials from other patients who have had similar health experiences.

ASSESSMENT AND PLAN:
Patient is nontoxic, afebrile with stable vital signs in the context of acute care visit as interpreted by me. No gross trauma or focal neurologic deficits on examination. Considerations include vasovagal syncope, dehydration, electrolyte derangement, rib fractures/contusion. Less likely cardiac syncope but will obtain screening evaluation. Patient was started on IV fluid and IV Toradol for supportive measures. Twelve-lead EKG is negative for acute ischemic changes or arrhythmia. Troponin is mildly elevated at 0.04 the setting of acute kidney injury. Repeat troponin was unchanged. Patient denies current chest pain. There is also mild hypokalemia and hypomagnesemia and oral repletion was provided. Chest x-ray is concerning for right perihilar and lower lobe opacity which is a change from 2 days ago. **Given concern for early developing pneumonia I will prescribe Augmentin and azithromycin. I suspect patient's presenting symptoms likely secondary to combination of vasovagal in the setting of current flu/developing pneumonia.** Patient is otherwise in no distress, nontoxic with stable vital signs and without new oxygen requirements. Patient will follow up with primary care provider after discharge for further evaluation. Strict return precaution was provided and patient expressed understanding.

From your provider:

"*I wouldn't worry about potential pneumonia just yet. Your symptoms could also just be indicative of the flu, syncope, or dehydration. Just to be safe, the medications I've prescribed you are used to treat bacterial infections such as pneumonia in particular.*

I will be checking in with the you in two weeks to make sure you're responding well to your medications. In the meantime, continue taking your meds as prescribed and be sure to notify us if you observe anything out of the ordinary."

Figure 10: Messages of comfort. An augmented assessment and plan section. When a patient selects a passage that was marked as having potential to concern them (highlighted in light blue), a side note appears. The side note offers reassurance from a clinician about why the passage should not necessarily concern them (if appropriate).