Collaboration, not Confrontation: Understanding General Practitioners' Attitudes Towards Natural Language and Text Automation in Clinical Practice

DAVID FRAILE NAVARRO, Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University
A. BAKI KOCABALLI, School of Computer Science, University of Technology Sydney
MARK DRAS, Department of Computing, Macquarie University
SHLOMO BERKOVSKY, Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University

General Practitioners are among the primary users and curators of textual electronic health records, highlighting the need for technologies supporting record access and administration. Recent advancements in natural language processing facilitate the development of clinical systems, automating some time-consuming record-keeping tasks. However, it remains unclear what automation tasks would benefit clinicians most, what features such automation should exhibit, and how clinicians will interact with the automation. We conducted semi-structured interviews with General Practitioners uncovering their views and attitudes toward text automation. The main emerging theme was doctor-AI collaboration, addressing a reciprocal clinician-technology relationship that does not threaten to substitute clinicians, but rather establishes a constructive synergistic relationship. Other themes included: (i) desired features for clinical text automation; (ii) concerns around clinical text automation; and (iii) the consultation of the future. Our findings will inform the design of future natural language processing systems, to be implemented in general practice.

29

CCS Concepts: • **Human-centered computing → User studies**; *Natural language interfaces*; *Text input*;

Additional Key Words and Phrases: Natural language processing, text automation, electronic health records, general practice

ACM Reference format:
David Fraile Navarro, A. Baki Kocaballi, Mark Dras, and Shlomo Berkovsky. 2023. Collaboration, not Confrontation: Understanding General Practitioners' Attitudes Towards Natural Language and Text Automation in Clinical Practice. *ACM Trans. Comput.-Hum. Interact.* 30, 2, Article 29 (April 2023), 34 pages.
https://doi.org/10.1145/3569893

Authors' addresses: D. Fraile Navarro (corresponding author) and S. Berkovsky, Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Level 6, 75 Talavera Road, North Ryde 2109, NSW, Australia; email: david.frailenavarro@hdr.mq.edu.au; A. B. Kocaballi, School of Computer Science, The University of Sydney, Building J12/1 Cleveland St, Camperdown 2006, NSW, Australia; M. Dras, School of Computing Macquarie University, 4 Research Park Drive, Macquarie Park 2113, NSW, Australia.

1073-0516/2023/04-ART29 $15.00
https://doi.org/10.1145/3569893

1 INTRODUCTION

General Practitioners (**GPs**) can be seen as the main curators of the patient's health record, especially in healthcare systems that rely on strong primary care. However, with the ongoing digitization of medical documents and **Electronic Health Records** (**EHRs**), this curation process has become a major burden [59] and a cause of burnout [130], so it currently represents one of the most time-consuming activities of clinician's practice [95]. Previous research has shown the importance of addressing stakeholders' needs in EHR development [19]. In the United States, and especially since the implantation of the 2010 Affordable Care Act [103], and its promotion of the Digitalization of Medical Records [39], it has prompted a great expansion of the Medical Scribe industry, inducing the appearance of increased needs towards clinical documentation.

Although human Medical Scribes have been partially addressing the growing needs of clinical documentation to a varying degree of success, their application continues to be rare outside the US and most of the documentation tasks still consume the clinician's time. Even in the US, there are growing concerns and calls to replace the medical scribes with automated solutions [29] and as such an interest in developing such solutions has been partially addressed by major players such as Google [13] or Microsoft [65].

One of the potential ways of addressing these issues relies on the use of automation and information technologies in Health Care which has a long and outstanding history [26, 36, 123]. Ranging from **decision support systems** (**DSS**) [118] to clinical imaging [43] and clinical prediction models [117], it spans a wide range of optimizations surrounding the clinical pipeline [112]. One of the most promising methods is the use of **natural language processing** (**NLP**) and **artificial intelligence** (**AI**) solutions, to automate clinical documentation and reduce clinicians' burden associated with manually entering information into the EHRs [93]. GPs are one of the specialities most heavily using free-text data, especially as they send or receive clinical textual information to communicate with other levels of care and specialities, interpret, and record diverse health correspondence, organize prescriptions, and so on.

With the recent advancements in machine learning methods, NLP has been showing promising results in several fields and application domains [37] including healthcare [101]. However, there remain some unanswered questions about the most beneficial use cases, the different ways clinicians could interact with text automation systems, and the fit of NLP-generated automated outputs to the clinical workflows. Preliminary work has been carried out, in areas such as generating an automated textual summary of clinical encounters [27], practical challenges faced by its development and translation [105], as well as clinicians' [66] and patients' [90] perceptions of such a system. Beyond technical limitations encountered in applying NLP methods to clinical text [21], there are additional challenges and questions around the most appropriate use cases for assisted automated technologies in real-world settings [126], especially in scenarios where high safety is required [128].

In this work, our aim was to investigate GPs' views on the capabilities of NLP and how automatic clinical text processing could support them in their everyday workflow. To this end, we conducted semi-structured interviews with 10 GPs actively working in primary care clinical settings in Australia. We identified that GPs generally have a positive view of text automation systems, pointing to a potential Doctor-AI collaboration. There are several key features that an automated clinical text system needs to perform, addressing various NLP tasks such as information extraction and summarization as well as certain aspects of the user interaction that need to be considered (such as GPs keeping control of the record, and creating searchable records), key attitudes of GPs towards text automation (allowing Doctor-AI collaboration) and addressing several concerns and challenges.

2 RELATED WORK

A wealth of literature has explored the attitudes toward different forms of automation [56, 108] in diverse, highly specialized and high-risk environments such as aviation [55], and banking [89]. However, past literature has covered more often the challenges of automation related to the use of mechanical machinery or robots [77] or concerning the use of traditional computational approaches and their interaction with human agents, for instance in finance [74]. Authors have also discussed "cyborg supervision" concerning other industries, such as banking [104], and the evolutionary rather than revolutionary approaches for human agents, as technology evolves towards more advanced forms of automation and inference [104]. More recently, attention has turned to exploring the recent developments related to the use of machine learning algorithms and more broadly AI [99]. Since the advent of the so-called "third rise of AI" [34], there has been an explosion of deep learning algorithms [44] capable of performing tasks that until recently, were believed to only be doable by humans. Progress in areas such as computer vision [63] and developments that spread beyond what humans are capable of, such as protein folding [61] or real-time weather prediction [106], demonstrate that AI systems performing cognitively complex tasks may become a reality soon. These newer developments bring to the fore the need for an increased understanding of the attitudes of users and how they would respond to the interaction with such advanced systems.

Although there is an increasing interest in the impacts of implementing advanced automation tools in the work environment, only a few studies have focused on the intersection between **Human-Computer Interaction** (**HCI**) and the use of NLP and text automation [80, 127]. Of note is that until relatively recently, the potential to automate highly complex text processing tasks was not materialised given the complexity and nuances of human written language and especially clinical free-text records [15]. It is only with the recent developments of sophisticated NLP algorithms [137] that automated clinical text processing has become feasible. In particular, the discovery of new deep learning language algorithms including statistical language models based on continuous vector representations of words [87] with subsequent development of transformer models, such as BERT [35], GPT-3 [14], and their specific derivatives, e.g., trained on clinical text corpora [2], have allowed their application for complex NLP tasks including named entity recognition [49], summarization [110], or question answering [51], also in the medical domain and the health industry.

The adoption of automated technologies in various aspects of the clinical pipeline is becoming more and more common over the years, especially concerning health information technologies where it can increase several quality indicators such as adherence to guidelines [22], increase surveillance and monitoring [22], as well as reduce medication errors. Studies have evaluated the use of various automated systems such as Decision Support Systems to guide prescription decisions [57] combined with earlier NLP developments integrated into the EHR [33]. Several authors have also risen questions surrounding the use of machine learning in clinical prediction models and the challenges this entails [23], especially given the potential for biases and the lack of transparency [92]. Similarly, previous literature shows the importance of tailoring algorithm predictions to user's needs, especially in the context of imperfect predictions [17].

More recently, clinicians have been asked about their views and attitudes toward AI and its application in various clinical specialities. An example is a dermatology, where there is generally an optimistic view of AI [99]. Other medical areas show conflicting evidence, such as radiology, where there is a discrepancy between the future expectations of the technology and the confidence of clinicians in AI-based results [62], although other studies have reported a more positive view [134]. General Practice, where the use of text records is pervasive, poses an area with both

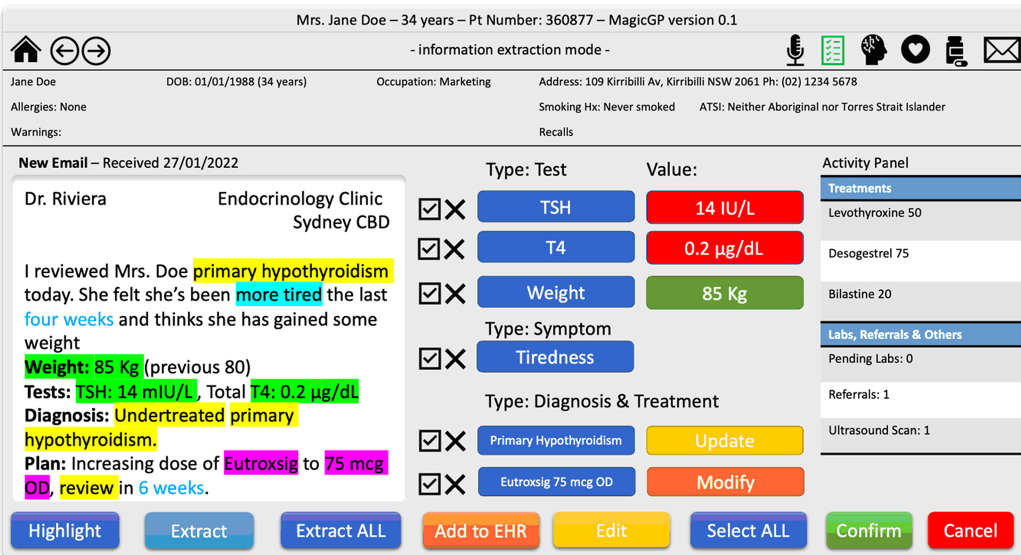

Fig. 1. Example of an automated Electronic Health Record System integrating Information Extraction components.

challenges and opportunities for automation [70], especially in the domain of NLP, as opposed to other specialities that rely more on imaging or other diagnostic technologies such as ECGs [7] or EEGs [97]. It was shown that GPs see potential in the use of AI to improve efficiency by reducing administrative burdens, but at the same time, several concerns remain, especially about the capacity of AI to provide clinical reasoning [9]. Concerning the specific use of automation technologies in the EHR, a few studies have addressed the attitudes of clinicians towards those, and in particular, showing the influence of the systems lacking optimization [59] as well as its influence on clinicians' burnout [119, 130]. However, to the best of our knowledge, there is little prior research on how NLP is envisaged to be used in certain scenarios and a particular clinical setting, such as General Practice.

The potential uses of NLP in medicine have been explored in previous reviews [138] as well as in the form of computational challenges addressing specific tasks [125]. One of the major related tasks that NLP practitioners have been actively developing is the information extraction task. This consists of building a system (either using rule-based methods, AI or a mix of both) that automatically tags certain words as recognisable entities. For instance, one of these challenges explored recognising drugs and their relation to dosage, administration route and reason for a prescription (disease it treats) [125]. Automatically unravelling this information from the free text would allow clinicians to accelerate and automate manual drug prescriptions, especially when those are being carried from one level of care to another, as exemplified by a fictitious user interface shown in Figure 1.

An important factor explored in the literature is that, to develop safe and successful AI-assisted clinical systems, ensuring system accuracy alone is not sufficient for the technology translation into a clinical setting. Aspects like explainability [3], automation bias [76], fairness [20], and other ethical challenges need to be addressed [129] in parallel with improving the precision and accuracy of AI. The added value of the automation, the integration into the clinical pipeline and, how the automated decisions are taken into consideration from a legal point of view are key considerations before any implementation may take place. Given the above issues, the need for user input starting from the early stages of the development process is crucial [66], especially as the particular context and non-linearity of primary care clinical consultations pose specific challenges for developing such systems [67].

Initial works on the attitudes and needs towards automated transcription [66] have shown the potential and interest of clinicians in using these technologies in clinical consultation, not only in primary care but potentially across the health system. The automation of medical documentation and EHR access, creation and maintenance has a major impact on the clinicians' burnout [119] as well as on resources and costs, given that currently, clinicians spend a considerable amount of time producing and consulting EHRs [95] even as there have been earlier calls for EHR reconceptualization [11]. This aspect has become particularly striking since the use of medical scribes has become a common "band-aid" solution to the usability challenges of EHRs [41] and in fact, has become an industry on its own [41]. An example of proposed approaches using NLP that led to implementation is described by Pine and Bossen [98] where clinical documentation integrity specialists use NLP in combination with document statistics for a computer-aided coding of the EHR. Another example is the Digital Scribe [27, 69]. As defined by Coiera et al. [27], it would employ a combination of speech recognition and natural language processing methods to automatically document the clinical encounter. Likewise, the preliminary studies by Li et al. [69] show a generally positive attitude of clinicians toward such systems, especially those that consider intermediate automation approaches. Although in previous literature, the focus has been on automating a particular part of the clinical encounter (e.g., transcribing the medical conversation into the EHR), they have not explored in detail the adequacy of the NLP tasks on their own (e.g., whether clinicians would value most verbatim transcription, summarization or extraction of specific named entities from conversations). Previous research has shown that the perceived usability of EHRs is an important factor influencing clinician burnout [85]. Given these factors, targeting primary care and the needs of general practitioners could inform the design needs and functionalities of text processing automation, in one of the areas of medicine with a stronger potential for adoption of AI, especially for improving time-consuming and low-value tasks [9].

Since first proposed by Weed in 1970 [135], the medical record has followed a problem-oriented approach. Subsequently structuring the consultation and its record followed the **Subjective-Objective-Assessment-Plan** (**SOAP**) mode [45]. This medical-record-keeping structure coexists and influences the traditional clinical consultation format described by Waitzkin [131]. However, consultations in primary care do not tend to follow these approaches in a linear way presenting additional challenges for automated clinical-text documentation approaches [67].

Research into human-automation interaction [60] and specifically AI has gained more attention recently [25]. Concerning the use of text automation, it has focused on various aspects such as using a human-in-the-loop framework [133] or aspects related to the user interface that facilitates data extraction [82]. In relation to health, a recent review explored the literature surrounding HCI in the mental health domain [122]. Previous work also has explored the design and implementation challenges of the EHR [115] as well as elements surrounding electronic document visualization [52]. Although doctors have previously made calls to reimagine EHR and design it with AI in mind [71] as well as considering certain elements of the EHR design and content structure such as the "problem's list" [141], still very limited literature exists on the design and key features that an NLP text system needs to provide; especially in the clinical-documentation context [66] and in primary care.

Hence, this study aims to (i) fill the existing gap in the literature on how text automation, in particular, related to the use of NLP technologies can be used and perceived by clinicians and (ii) explore the needs of clinicians, their attitudes, and the desired features of text automation systems deployed in General Practice. Our work has the potential to guide future clinician-technology interaction studies and facilitate further developments of AI-driven text processing tools for clinicians.

3 METHODS

To develop products or systems and explore user preferences, interviews with potential users are described as one of the key steps to understanding their needs, attitudes and opportunities for the technology [50]. Interviewing individuals to uncover their values and preferences, before further refinements such as live prototyping [53], or other more specific techniques to explore interactions, such as Wizard of Oz [136] can facilitate testing the more relevant ideas and design elements.

We conducted semi-structured interviews[1] followed by a thematic analysis of the interview transcripts to find the main themes and sub-themes. Semi-structured interviews offer specific advantages and have been used extensively in qualitative research, specifically when interviewing health professionals [31]. Inductive thematic analysis is commonly used for synthesising and analysing ethnographic data [12], including interviews [38].

3.1 The Interview

In the introductory part of the interviews, the participants were provided with a brief introduction to the technology. This consisted of an oral presentation by the interviewer that outlined the meaning of the technologies, aiming at bridging any prior knowledge gaps. This introduction clarified the following concepts: automation, Artificial Intelligence, Natural Language Processing, free-text, voice-recognition, and text-to-speech technologies. Additional clarification and explanations were provided on specific NLP tasks that were discussed: information extraction [5], named entity recognition [91] and summarization [81]. The script of the initial explanatory pitch is provided in **Appendix 1**.

Following the introductory part, the interview itself commenced. The interview questions were semi-structured, where the interviewer ensured that the conversation covered all the prespecified topics targeted by the interview questions. We piloted the questions with the clinical academic health practitioners at Macquarie University, who helped to refine them.

The questions were structured in two thematic groups. The first was aimed at exploring the general attitudes of the participants toward AI and the use of automation technologies and related techniques in their day-to-day routines. The second group of questions related to NLP technologies in particular and their use in specific clinical and documentation scenarios. The guiding questions are provided in **Appendix 2**.

If the GP could not propose plausible documentation scenarios on their own, they were presented with examples of clinical scenarios helping facilitate their thinking process. These were based on the authors' clinical expertise and exemplified typical cases that may reflect GPs' documentation processes. In these scenarios, participants were prompted with a typical consultation situation: reviewing documentation, such as a newly generated hospital discharge summary, recording information during a clinical visit, such as recording a new diagnosis or prescribing a new treatment. The full script for these scenarios is provided in Box 1.

The last question was rather exploratory and proposed to the clinician to envision the GP practice of the future and imagine a new consultation paradigm that integrated the elements and technologies discussed, as well as any other element that was not specifically covered by the interviewer and the questions.

[1] The study was approved by the Macquarie University Research Ethics Committee, reference 52021931324252.

Box 1: Clinical Documentation Scenarios

Scenario 1:
OK, imagine now the following scenario. You have a patient that is coming today to see you, he has recently been discharged from hospital after a complication of his chronic disease. You have received early this morning the discharge letter fromthe hospital, where it explains the process, the patient has followed, the test they performed, and inform you of new diagnoses and changes in his medication. The patient is coming soon, and you have quite a loaded morning, so you were not able to read or do anything with this discharge letter beforehand.

In this scenario, how do you think text automation could help you?

Scenario 2:

There is a new patient coming to visit you. She is moving from a different practice is an elderly woman with a long medical history, past treatments, and various illnesses. You have not previously interacted with her, nor do you know what her reason is to visit you. Your current EHR is unpopulated besides the basic patient information.

In this scenario, how do you think text automation could help you?

3.2 Study Setting and Participants

Previous literature on qualitative research has shown that for investigating phenomena surrounding experience, a sample size of six participants is typically sufficient [109]. Other sources [46] suggest that a sample size of 12 thematic saturation is more likely to happen. We recruited participants until reaching thematic saturation, which already happened with a sample of 10. To calculate saturation, we performed a retrospective analysis to assess data saturation following the methodology of Guest et al. [47]. Using this approach, a set of the first six base interviews is initially assessed regarding the amount of new information they create. The ratio of unique codes produced is used to estimate the quantity of new information that would be produced in subsequent interviews. This number is then used as the denominator to calculate the ratio of new information (unique codes) created in each subsequent interview. By calculating the ratio of new information for the remaining interviews, each interview that produces less than a 5% threshold of new information can be identified, and therefore, evaluate after which interview saturation was reached.

Australian General Practitioners, currently practising clinical medicine, at least part-time and with experience using any Electronic Health Record system were deemed eligible for participation. To simplify recruitment, we did not set any limits on the age, location, or working experience of the GPs. Participants were recruited through various channels. We created an invitation leaflet that was sent through the university channels. We also reached out to the local health district, which distributed it further among local GPs and practices. Lastly, we completed the recruitment through snowball sampling after the initial interviews. All the participants provided informed consent to collect and utilize their answers and publish the results in peer-reviewed publications. Participants were compensated for their time with an e-voucher of A$150.

We interviewed 10 GPs from 3 states in Australia (New South Wales, Victoria, and Queensland). All the GPs interviewed currently practise, at least part-time, in clinical practice with a varying range of roles and interests (academic, commercial, entrepreneurial, technological, and more). Out of the 10 included GPs, 9 worked in urban areas, 1 worked in a rural setting and 30% were women. The GPs had all over 10 years of medical experience and more than 5 years of working in General

Table 1. Participant's Characteristics

Participant Identifier	Years practising medicine	Years as GP	Practice Size	Practice Location
Participant 1, Male	10	8	1	Urban
Participant 2, Male	15	15	20	Urban
Participant 3, Male	17	10	11	Urban
Participant 4, Male	10	5	10	Urban
Participant 5, Male	12	8	9	Rural
Participant 6, Male	14	10	12	Urban
Participant 7, Male	11	9	17	Urban
Participant 8, Female,	14	9	5	Urban
Participant 9, Female	10	6	20	Urban
Participant 10, Female	33	29	16	Urban

Practice (5–29). The size of their current practice, at the time of the interview, varied from 1 to 20 GPs, with 6 of them working in practices with more than 10 GPs. All the GP included in the study used EHRs in their day-to-day practice, and none of them had any previous experience with the use of text automation technologies in EHRs. Table 1 summarizes the demographics of the 10 included GPs.

3.3 Data Collection and Processing

Given the remote locations of most of the participants and limitations due to the COVID-19 pandemic, 9 out of 10 interviews were conducted through the Zoom video-conferencing platform and only one interview was conducted in person. The results were stored on the local researcher's computer in compliance with the University data management policies. The interviews were video-recorded and transcribed using a commercial transcription service. In the transcription process, any personal identification was removed from the transcripts. If there was a specific person, group or organization mentioned in the transcripts, they were given pseudonyms.

The participants' responses provided us with a rich data source on how GPs envisioned the use of automated text processing technologies, how they see them integrated into clinical practice, and further information on interacting with technology, safety and privacy concerns, as well as the GPs views on the clinical practice of the future with a high degree of automation. We followed a bottom-up, thematic analysis [12] approach to extract the main themes and sub-themes of the conversations. The initial coding of a sample of the interviews was performed by 2 researchers, using the NVivo® software.

The thematic analysis was structured in six phases: (1) completing the review of recordings, transcripts and notes; (2) individual coding to allow a diversity of coding approaches to be discussed, combined and reworked to generate codes; (3) discerning the emerged themes and subthemes from the patterns present in the codes; (4) reviewing the emerging candidate themes and subthemes then refined by examining them in relation to the dataset; (5) establishing and defining the themes and subthemes that capture the best the final concepts; and (6) results reporting [121]. To develop the initial codebook researchers coded in duplicate a sub-sample of 20% of the interviews to unify the criteria for developing the subsequent codes. After this initial adjustment, the complete set of interviews was analysed, and the final codebook was defined by consensus between two researchers (DFN, ABK). Disagreements were resolved with the help of a third researcher (SB). Once the coding of all the interviews was completed, the generated were consolidated into 18 categories through discussion. Further discussion between researchers led to the compilation of the emerged unified themes and subthemes that were decided by consensus (between DFN and ABK), with the help of a third researcher (SB) in case of disagreements.

Conducting a retrospective analysis, we observed that the first six interviews in this study produced 18 unique codes. We calculated the percentage of new information in the remaining four interviews of the initial sample. It was found that the four additional interviews did not reach the 5% threshold for new information, as they did not produce new codes and therefore, data saturation was reached sufficiently with the initial sample of 10 participants.

4 RESULTS

When GPs were asked initially about their attitudes and views towards the use of text automation technologies, they expressed a keen interest in these technologies and generally a positive view regarding their potential benefits:

> *I think they have the potential to be incredibly useful. Particularly I guess where I'm thinking you're coming from is in relation to medical records. So one of the problems I find as a GP and what a lot of colleagues tell me as GPs. Is that they struggle to see patients and then record their medical notes after the consultation, within the 15 minutes that they have. So any technology that allows GPs to more accurately record records in a shorter period of time would be incredibly useful. (P 3)*

> *It will make life easier, and it will integrate and maybe improve care for the patient, that is what I'm thinking. (P 2)*

> *It sounds like it could be very positive. I don't have too many reservations because I feel like we already store all of our data in an electronic format. (P 5)*

Four main themes surrounding text automation technologies in general practice emerged. We note that the GPs expressed a generally positive attitude towards what can be framed as doctor-AI collaboration. They discussed several desired features, although also raised some concerns and challenges the system design and development need to address to pave the way to successful implementations in the consultation of the future. Table 2 presents a classification and description of the emerged themes and sub-themes.

Table 3 shows how the Desired Features and Consultation of the Future themes and their subthemes fit into consultation tasks at various parts of the consultation using specific NLP technologies (last column). Pre-consultation indicates activities that are performed by the clinician before the patient is seen. Consultation indicates activities that take place at the time of the patient's visit, while post-consultation refers to activities completed after the end of the visit. "NLP Tasks Required" describe the NLP technologies that can facilitate achieving each of the above features. Text-to-Speech recognition encompasses the technologies that are employed to transform voice into text. Named Entity Recognition refers to technologies that allow identifying and extracting specific entities (e.g., drugs or diseases) from a free text. Relation Extraction indicates those NLP methods and systems that establish relations between the extracted entities. Lastly, Summarization refers to the technical systems that can generate concise summaries of long texts (e.g., conversations).

4.1 Doctor-AI Collaboration

One major theme that emerged in the interviews was that GPs perceived more opportunities to collaborate with AI than reasons to distrust these technologies. The key element of this collaboration resides in defining how to establish a sensible relationship between doctors and the technologies

Table 2. Description of Themes and Subthemes

THEMES	SUBTHEMES	DESCRIPTION
DOCTOR-AI COLLABORATION	(1) Progressive Automation (2) Keeping control of the record (3) Explainable & transparent AI (4) Trustful AI	Elements that define the relationship between Doctors and AI technologies.
DESIRED FEATURES	(1) Extracting information (2) Digital Scribe (3) Creating summaries (4) Searchable records	Useful functionalities of an automated clinical text system.
CONCERNS AND CHALLENGES	(1) Resistance to change (2) Medico-legal issues (3) Privacy and safety, (4) Automation bias (5) Filtering and aggregation (6) Implementation	Major barriers and areas of uncertainty around text automation technologies.
CONSULTATION OF THE FUTURE	(1) Beyond screen and keyboard (2) AI as an expert system (3) AI as a digital assistant	The vision of practice with new AI-enabled features.

to avoid its pitfalls. Notably, Doctor's attitudes towards automation technologies did not convey opposition to them, as a threat of AI potentially replacing doctors. This concern was residual in the interviews:

> *I know that people* talk *about AI replacing clinicians. I don't think that's where we are at the moment. I still think that clinical acumen, that knowledge, that contextualizing it for the patient is quite important. But the technologies themselves, [...] it would actually make you quite efficient, make the working day more enjoyable that you have more time to spend with the patient and to contribute to patient care. (P 9)*

We identified four subthemes within doctor-AI collaboration:

4.1.1 Progressive Automation. A few participants indicated that a progressive implementation of automation technologies has a higher likelihood of being accepted and implemented. The process of transitioning to automated text processing should not be accomplished at once, but the system should instead take over the tasks gradually, requiring less and less input and control from the user over time.

> *I guess over time as it improves, it's more precise, fewer errors, and there's more trust in the system, then yes, a lot more. . . Maybe there could be a bit less checks and balances. (P1)*

> *If there are technologies developed, and you're first working with them, you probably do want to have that control. (P 10)*

4.1.2 Keeping Control of the Record. An important collaboration feature that some participants wanted to preserve with text automation was having control of what goes into or out of the medical records. So even if automated text analysis or voice transcription is implemented, there should still be a "veto right" for doctors.

> Doctor *should have control over whether or not the whole conversation is recorded and stored. Or whether their conversation is recorded, stored, processed, and then deleted.*

Table 3. Desired Features and Consultation of the Future: Aspects Subject to Automation in the Consultation [135] and Medical-record [131] Processes and Respective NLP Tasks that can Facilitate these Features

Consultation component and when it occurs (Pre / Consultation / Post)	**Subthemes** *Italic "Desired Features"* **Bold "Consultation of the Future"**	Text-to-speech recognition	Named Entity Recognition	Relation Extraction	Summarization
History Taking (Consultation, Post)	*Digital Scribe*	●	●	●	●
	Creating Summaries				●
	Extracting Information		●	●	
Physical Examination (Consultation)	**Beyond screen and keyboard**	●	●	●	●
	Digital Scribe	●	●	●	●
Results Review (Pre, Consultation)	**AI as an expert system**		●	●	
	Extracting Information		●	●	
	Creating Summaries				●
	Searchable Records		●	●	
Assessment (Pre, Consultation, Post)	**AI as a digital assistant**	●	●	●	●
	AI as an expert system		●	●	
	Searchable Records		●	●	
	Beyond screen and keyboard	●	●	●	●
	Creating Summaries				●
Treatment (Consultation, Post)	*Extracting Information*		●	●	
	AI as an expert system		●	●	
Referrals, Requests & paperwork (Consultation, Post)	*Extracting Information*		●	●	
	AI as a digital assistant	●	●	●	●
	Creating Summaries				●

And just whatever the doctor approves to go into the EMR (Electronic Medical Record) is what remains. (P 1)

As long as I could give the final yes or no. (P 5)

Flag it for me that this is different, is it something that I want to add or is it not. Because making me aware of it firstly is a really good step and then anything extra is a bonus. (P 7)

Reasons for keeping control of the record are varied but generally involve Doctors being seen as the curators of the records and preferring to maintain a subjective and highly personalised version of the patient's history. They consider this role should still be maintained even with the increased automation, e.g., for automatic recording and summary creation, and text extraction.

As long as it's the same system where I can accept it and then further edit it. Because I'm quite specific when I add to the past history. (P 5)

However, although there is a desire of keeping control of the record, the trade-off between having full control and potentially less useful automation, and not having enough control but having a potentially smarter system was also mentioned.

> *I suppose it is that balance isn't it, I still want to have control over some things, and it's because I know that the sort of information that comes into our system, there just isn't enough information there for it to be able to happen automatically. (P 7)*

> *To review what's been done, but that defies the point, because what's the point of having AI if you still have to manually go through the whole document? (P 8)*

4.1.3 Explainable/Transparent AI. A few participants understood the explainability trade-offs and limitations of AI-based technologies, especially relevant when automated decisions and algorithms are exploited. They would appreciate it, if the system informed them of the different options and likelihood of specific clinical items, such as a disease, a symptom, or a drug, especially if they were integrated into a decision support system or performed inferential tasks such as suggesting a diagnosis.

> *Instead of saying "we think this is pancreatitis" or "pancreatitis is the most likely diagnosis" I think what I think would be better would be just a kind of, "before you order this test, these were the differential diagnoses and pancreatitis was one of them. Now that you've done these tests, of those five differential diagnoses, only pancreatitis remains". (P 3).*

Another important element was that automated systems would need to provide sufficient transparency on how the data is being processed and how secure that processing is.

> *Part of it would be an explanation of how it works from a privacy level. What is listened by the computer during a consult? How that information is stored. Is it stored locally? Does it get sent to the Cloud? Is it processed in the Cloud, or is it processed on a local machine? And how secure is that? Is it encrypted?. (P 1)*

4.1.4 Trusting AI. Although participants are wary of the possibility of AI misjudging, an important enabler for successful and productive doctor-AI collaboration is building trust in the system. This would allow a successful implementation of the technology in clinical practice.

> *Once you trust the system, if it could just automatically do all of that [extracting information from discharge summaries], would be great. So, add new medications, add classifications or diagnoses (P 4).*

> *I wouldn't use it blindly I guess, and so I'd be confident that it would save me time but not replace me thinking, which is not the aim, for me it's the saving time. (P 7)*

> *You'd need to feel confident it was accurate, and that information was reliable, that it was able to properly look through, for example, if it's a bunch of old notes and things, that it actually was checking everything accurately and extracting the right stuff. (P 10)*

The theme of Doctor-AI collaboration features several important subthemes that define the relationship between the clinicians and automated AI technologies. The key themes raised in our interviews include the progressive implementation of AI technologies, the ability of the clinicians to keep control of the medical records, the explainability and transparency that AI should exhibit, and trustful relationships that need to be established between the AI tools and the clinicians using these.

4.2 Desired Features

Another important theme that emerged through the conversations was a set of desired features that the automated technology needs to offer, which are important for clinicians in their day-to-day clinical tasks. These tasks are categorized below following the structure of the existing taxonomy of NLP tasks and subtasks [68] and the proposed modes of interaction with such systems.

4.2.1 Extracting Clinical Information (Information and Relation Extraction). When clinicians were prompted with questions surrounding the processing of already written text (e.g., correspondence, discharge letters, medication reviews, referral letters) they opined that being able to extract specific information from these texts would be a necessary feature.

> *And hopefully detecting the keywords, like particular symptoms, and signs, and duration, and quality, and characteristics of what the patient's telling. And then translating that into a succinct, concise medical note that is similar to what a doctor would write for their notes. (P 1)*

> *If there were quick ways of extracting some key things, like key diagnoses, the medications, who their specialists are, and that's then in your notes, … That would be helpful. (P 10)*

An important aspect of the information extraction tasks is how the Doctors envision it fitting into their interactions with the EHR. For this reason, detecting clinical entities and highlighting them in the clinical text, i.e., named entity recognition and relation extraction, would be an important feature.

> *The most useful application of that would be for the natural language processing to read the discharge summary. And then suggest what could be updated in the patient's file. If the natural language processing may pick up a dose change, or a change in medication, then it would suggest updating in the EMR or electronic medical record. And the doctor would just approve it, or edit it by… To make it faster that way. (P 1)*

Additionally, the system should be able to detect discrepancies when comparing the free-text records from different sources (e.g., from a discharge letter from the hospital) and the local electronic health record (the primary care EHR):

> *It would be very useful to compare the text to what's in the file and make adjustments or have options to make easy adjustments. Entering medications is actually quite a cumbersome job, and so anything that could help with that would also be a valuable thing. (P 7)*

4.2.2 Digital Scribe – Automated Voice Documentation (Speech Recognition). Another important feature for participants was the ability of the system to capture the clinical conversation and transcribe it directly into the medical record. Having this feature would allow them not to type during the patient's clinical visit or afterwards. This feature could improve substantially their interactions with patients, as it will allow clinicians to be released from typing as they see patients and focus instead on clinical questions and examinations.

> *At the moment we use a system called MedicalDirector [143] where we need to type in most of the notes. If you're talking about an actual language processing system,*

where when we are talking it starts to record and then translate the consultation into clinical notes, I think that will be useful, If there were a program that could record my notes without me needing to sit and do it afterwards, that would save me time. (P 2)

The summary of the consultation, it could certainly help, it could certainly be useful there. (P 7)

4.2.3 Creating Record Summaries (Summarization). Another characteristic that participants found valuable was the ability to generate summaries of the whole patient record. This feature would prove especially useful in scenarios like a new patient coming to the practice or a doctor that practices at multiple sites, with unfamiliar patients.

If you requested all the notes from their previous GP, for example, and then I guess you wanted the system to analyse, and create a brief summary, or transcribe that information into the electronic medical record. I think the greatest use would be to… Like for new patients, their notes from their previous GP to naturally transcribe that into your electronic medical record. (P 1)

I think it'd be very helpful if it could provide a clinical handover summary of the patient almost with the key person and features that you're normally looking for. (P 9)

Another desired feature was the possibility of aggregating historical patient data into a coherent summary. This feature would imply using information extraction and summarization techniques across multiple documents to provide a brief resume of the most relevant diagnoses, episodes, treatments, and elements in the entire patient record.

For example, the doctor in the previous consult records this significant problem and then you say, "Okay, can you bring up some significant past medical history?" Then you should be able to filter out because most notes are quite huge, it will filter out and bring up the important things. And that probably will be useful before you start to see the patient. Maybe there should be some way for a doctor to mark this as important so that it brings it up and the AI can bring it up later on when another doctor looks at the notes. (P 2)

Additionally, another important feature that was highlighted is the ability to generate summaries from the free text record, that can be transcribed into a clinical letter or referral letter.

One of the things that people want to look at also is to be able to summarise a consultation, which is held in natural language, into a summary letter without anyone needing to type anything. (P 4)

An interesting feature participants pointed out was the possibility of providing in real-time a plain-language summary for patients. Thus, after the clinical visit, the system would not only produce a relevant summary to be included in the EHR but simultaneously generate a different summary that the patient could receive and take home. The latter would summarize the consultation and the treatment options in simple words.

I thought might be actually really helpful if, at the end of a consultation it[the system] automatically sends the patient a summary of the plan. I guess one of the things that you often tell patients is you talk to them about. . . If they've got a cold and if you

> *don't think it's COVID, we might say, try some warm steam, some days we'll say, ... if that just gets automatically emailed to the patient, that will be amazing. (P 4)*

4.2.4 Searchable Records. In conjunction with extracting information from free-text records, making the medical records searchable (e.g., providing a search engine interface like Google) was deemed valuable.

> *GPs receive discharge summaries or letters from specialists as like PDF documents. So it's not really possible to search those documents for key terms or keywords. I'm wondering whether there may be some opportunity to be able to search documents. Search for things within documents and so then that can help if a patient comes in with an undifferentiated presentation, to be able to better understand what has previously occurred (P 3)*

> *And also, something that I didn't mention earlier is also if there's any way to quickly search or locate information. Because there are times like, I don't even know, a patient comes in and is it must be in the notes. (P 6)*

The Desired Features theme summarizes the functionalities that are perceived as important by clinicians when envisaging the use of text automation technologies in their practice. Several themes have been brought up, revolving around effective information extracting, deployment of automated Digital Scribes, the ability to create consultation summaries, as well as the ability to easily search through the records. It should be mentioned that many of the above have been subject to extensive research in Natural Language Processing and Information Retrieval.

4.3 Concerns and challenges

Although participants were aware of the advantages that text automation may bring into their day-to-day practice, there were a few concerns and challenges that clinicians identified that could reduce the trust in the system, hinder adoption, or pose issues for both clinicians and patients.

4.3.1 Clinicians' Resistance to Change. Participants considered that some of their colleagues and organisations may be resistant to implementing changes to the consultation or the record system. This could potentially limit the pace of implementation and adoption of the technology or could make it necessary to keep both automated and manual systems in place.

> *I think that general practice obviously, in the medical field in general we're quite behind with trying new technologies in clinical practice. Bear in mind that I probably have a different view compared to other GPs perhaps, because I am in the entrepreneurial and health tech start-up space. I'm not sure if all doctors would feel that way. And I come from a different demographic, obviously. A younger demographic, maybe, compared to the 50-year-old GPs out there. (P 1)*

> *Doctors more towards near retirement, they might not be happy to take up the new technology. ... it comes back o the medical-legal aspect. (P 6)*

4.3.2 Medico-legal Issues. One of the main concerns was the legal implications of text automation, e.g., in case some clinical information is misused. Participants were worried that if the patient mentioned something that was not considered and there was a verbatim record of everything that happened in the consultation, this might be used against them in legal procedures.

> *If everything was completely recorded and stored, [...] the entire conversation, if that was recorded and stored, and the doctor missed something, then that could be used*

> *against the doctor. That the patient mentioned it, but it wasn't listed or written, or a doctor accidentally missed that detail. Then that could be a problem. (P 1)*

Another problem may appear if some clinical information is incorrectly transcribed, or if the system misses it, then errors may be carried over multiple systems and through multiple patient records.

> *If, for example, they wrote schizophrenia instead of schizotypal personality disorder. Somebody then has a diagnosis that maybe was incorrect because somebody clicked on the wrong button. Then everybody else, the GP, then the next GP and then the other mental health service might then label this person with a condition that they don't have. (P 2)*

4.3.3 Automation Bias. Participants perceived a potential risk that may emerge with the use of automation tools not being able to conduct tasks that would have been otherwise done by a human doctor, such as correcting obvious errors. Previous work has described this type of error as automation bias [75].

> *Perhaps a medication error, somebody makes a mistake on the discharge summary about the medication that was dispensed on discharge [...] if you're doing it manually, might think that dose doesn't sound correct. Whereas if it's done automatically the GP might not think about what's been copied and so the error then gets multiplied. (P 3)*

Additionally, it may also make clinicians less attentive to reading and processing complete information, especially if certain parts of the record naturally become more salient than others.

> *When It [text automation system] emphasises things, sometimes it might emphasise one part and then the doctor automatically doesn't look at the other parts. [...] Then you might get an issue where the doctor using the summarised documents, actions only part of the recommendations from the hospital. (P 4)*

> *I can see potential for distraction, but I can also see potential for glossing over of information. For example, if I got a patient to free input a whole lot of data before I saw them, and they said no to some questions and yes to others, then I probably would be more inclined not to ask them the exact same questions again and be focused more perhaps on the stuff they've said yes to. (P 5)*

4.3.4 Privacy and Safety. Participants raised concerns regarding the privacy of the system, the data it stores, the text processing functionality, and how the confidentiality of the patient's personal health information was maintained.

> *I'm not sure about confidentiality, whether anyone who looks at the notes needs to know everything that is going on or not. I'm not sure how we can manage that. Though, I don't think that will be a problem because I think once the patient is coming to see you, it means that you have access to the whole set of notes. (P 3)*

> *It's understandable because it's people's medical information. (P 8)*

Equally, concerns about the safety of the system were raised, for example, when using automatically recognized entities to prescribe new treatments or codify allergies.

> *And also, in case it recognizes the discharge summary. I guess there's a lot of difference between the patient does not have a penicillin allergy versus the patient developed a penicillin allergy. (P 4)*

It also may pose a risk of misidentification and incorrect labelling of patients, medications, and diseases.

> *It might misrecognise a mistyped medication. Instead of, say, prednisone, it takes prednisolone. It's a minor error. But you know what I mean, it accidentally puts in the wrong medication because of, it could be the errors in someone else's letter. (P 4)*

4.3.5 Filtering and Aggregation Challenges. Participants had concerns regarding the system's ability on differentiating relevant and irrelevant data, as this may create either big and unmanageable records or may cause clinicians to lose important pieces of information.

> *Maybe it may show up too much information too if it's not able to filter out what is important, not important, it may show a huge list of problems there, which is based on symptoms rather than diagnosis. So that could be a problem, too. (P 2)*

> *Transcription as in verbatim transcription will not be very helpful because it's just too cumbersome [...] But if there was a technology that can get the key information out of a conversation. (P 7)*

Equally, this concern was raised in relation to creating automated summaries, or transcriptions of conversations and how capable the system would be in distinguishing what is important, from a clinical point of view, from what is not and in capturing clinicians' judgments and subjectivity automatically.

> *When anyone writes their notes ... there's a degree of subjectivity in there. As a clinician, I'm really deciding what I feel are the key points that are worthwhile to document. I question how well a computer program could identify the same key points that I identify. (P 5)*

> *There's a lot of conversation that you don't necessarily need in the notes, so it's got to be selective, doesn't it?. (P 10)*

4.3.6 Implementation. Another challenge mentioned by participants referred to the need for the system to be integrated with current EHRs and information technology infrastructures. The question related particularly to the ways the current governance of EHR systems or software providers may limit the use and implementation of a new technology that may require considerable changes to the existing systems.

> *When I was trying to look into similar thing I guess there was an issue with these technologies. So, unless you make the own EMR clinical software, or you have some kind of agreement. (P 6)*

> *It would need to obviously work well, it would need to integrate really easily with our current clinical practice software, and there are multiple. (P 7)*

However, participants considered that this barrier could be overcome if there was a willingness to change the ecosystem.

> *I know, for example, Software 1 doesn't allow a third-party software to alter medical records at this stage. In terms of adding in new conditions, or removing them, or changing the dates, or anything like that. (P 1)*

The Concerns and Challenges theme discusses the major barriers and areas of uncertainty around the use of text automation technologies. Several important challenges were raised touching upon the clinicians' resistance, potential medico-legal issues, technical questions surrounding privacy and safety, risks of automation bias, and practical implementation challenges. It is evident that some of these challenges are critical and need to be addressed in future research, within and beyond the computing disciplines.

4.4 Consultation of the Future and New Ways of Interaction

Participants envisioned that in the future the consultation pipeline might become different from what they are used to nowadays. The use of a combination of new technologies, including virtual and augmented reality, voice recognition, conversational agents, and voice assistants, may lead to a more patient-centred consultation than in today's highly computer-centred setting.

4.4.1 Beyond Screen and Keyboard. A common theme was the idea of moving beyond the screen in contrast to the current computer-centred records and interactions, supported by innovative technologies such as augmented reality.

> *Maybe in the future with augmented reality you could see more visually maybe an overlay over a patient or like in space[...] You'd be talking to a patient, in the top left it would say, please note, this patient has a history of COPD. Has these many exacerbations per year. High chance of hospitalization or something like that. (P 1)*

> *I suppose in an ideal world, I would have minimal hands-on input into the computer. But it would record my interactions and perform the functions that I need to function seamlessly, quickly, accurately. (P 7)*

Another aspect of interaction that participants envisioned was voice command, as opposed to the current primarily mouse and keyboard-based input.

> *It may be useful to say "can you bring up the discharge summary from so and so date or for so and so admission". I think that will be useful. (P 2)*

> *Yes, just taking my hands off the computer, getting my eyes off the screen, so that I can be spending time with the patient. And also saving me the documentation time, because you can either spend more time with the patient or see more patients. (P 7)*

4.4.2 AI as an Expert System. Participants envisioned AI as a virtual expert sitting nearby, that could prompt potential relations and associations arising from the integration of free-text notes and other medical data, such as lab results, into the decision support processes.

> *It's maybe like a second opinion right there. Maybe imagine you have the professor of general practice sitting there with you, guiding you in your consult. What an amazing experience that would be for both the doctor and the patient. (P 1)*

> *If it flags it to you that there's an abnormal finding or it takes the next step and it actually calls back the patient, the doctor needs to see you for a review because there is a finding in the chest x-ray. (P 9)*

4.4.3 AI as a Digital Assistant. Participants could also see that automation might take away from clinicians some mundane tasks such as managing reminders or scheduling appointments.

> *It wouldn't be like imagine, like, in five years or ten years' time, it will be like, before the patient enters, like "show me a summary" or "show me significant medical issues" or "what did we deal with in the last medical consultation?" or "what changes did we do" Something like that. And then once the patient comes in, maybe it should take notes automatically without me needing to type it. (P 2)*

> *At the time of booking it asks the patient why they've made the appointment. So perhaps there's some key symptoms or questions that it might ask [...] to try and ascertain, to get more information. (P 3)*

Another important feature envisioned was the possibility of speech-to-text generation when producing clinical text, similar to smartphone predictive keyboards and text auto-completion systems.

> *In terms of note-taking I wasn't sure when I sign up for your study, when you meant auto texting. So, I know like when I try to email on Gmail, they recognise what sentence I might want to write and give you possible text. So even that will speed up the process. (P 6)*

Lastly, another element highlighted was automatically sorting correspondence, e.g., email categorization and prioritization.

> *I guess it's an idea that I've had to be able to automate the automatic cataloguing of correspondence that comes through in general practice. For example, exactly as you're saying, discharge summaries, specialist letters, lab results, if it comes in, it's automatically categorized because there are some letters that are pretty much useless. (P 4)*

The Consultation of the Future theme refers to the ways the clinicians envisage AI being deployed and interacted with in the future. Intriguing ideas have been raised, including moving beyond the current screen-and-keyboard paradigm, and using AI as an Expert System and a Digital Assistant. While some of these are still seen as futuristic in health care, somewhat related technologies have matured in other applications, e.g., chatbots or shopping assistants, giving hope to similar technologies being applied in medicine.

5 DISCUSSION

Our findings suggest that GPs generally have a positive attitude toward text automation, and they would be open to NLP-based systems being integrated into the General Practice. Although several concerns were raised, none of the participants indicated they would object to using these technologies, if properly tested, developed, and implemented.

We identified main themes describing the attitudes of GPs towards automation. These themes are connected through the overarching concept of Doctor-AI collaboration, where a trusting relationship is established between the clinician and the automated system. There are several important design considerations, and features to facilitate meaningful collaboration. The main ones are around designing systems that allow information extraction, voice recognition and meaningful summarization, and successfully integrating them into a new EHR going beyond current paradigms of user interaction. There are challenges as well, that need to be overcome and that may limit the chances of a successful implementation and adoption, the main ones being addressing medico-legal issues and interoperability with current systems. Lastly, clinicians identified the opportunities for new ways of interaction and envisioned a not-so-distant future, where clinical consultation documentation moves the screen-mouse-keyboard paradigm, with AI systems taking over the automation of the tedious clinical documentation tasks, allowing doctors to spend more time on patient care.

Our findings are aligned with previous studies that have shown the emergence of the narrative of collaboration between the user and technology (and specifically AI) and its importance for system designers [18, 123, 132]. Cai et al. [17] highlighted the concept of human-AI collaboration, but equally stressed the need for upfront information and the strength and limitations of the AI models to be used, similar to our findings regarding clinicians' desire for a transparent AI. Wang et al. [132] explored the subject of Human AI collaboration in relation to its HCI components in medicine, and in particular in the case of rural GPs who, despite some concerns, expressed a generally positive attitude toward AI decision support systems [133]. Human-AI collaboration has appeared in relation to GPs and the use of automated documentation assistants [66], where the idea of the clinicians keeping control of the record was also expressed by maintaining a supervisory role over the AI agent. In contrast to our study, Buck et al. [16] and Kocaballi et al. [66] found a certain degree of existential anxiety in GPs concerning the use of AI-assisted diagnostic tools, as our study surfaced no major concerns surrounding the replacement of doctors in the near future.

In our interviews, this human-AI collaboration was mostly imagined as a hierarchical relation, where the human agent takes a supervisory role overseeing the work of the AI system, which acts more like an assistant. However, doctors also leave space for further developments, where successive waves of progressive automation may allow the text automation AI to gain progressive independence from the human agent, increasingly acting like an expert. An interesting and useful idea for the doctor-AI collaboration was suggested by a prior co-design study with GPs [66]. According to that study, doctor-AI collaboration may work like a pilot-autopilot collaboration model, where both the actors play a critical role in the successful completion of a flight: "On the one hand … pilots act as a safety guard. On the other hand, autopilots were considered essential assistants to help control the aircraft during the entire flight." Previous research has also shown that the use of a human supervisor as an "arbitrator" among several ML models might be a successful approach, especially when implementing AI in settings requiring life-critical decisions [40] or when deploying these systems in specific settings such as in rural context [133]. Equally, it has also shown the importance of addressing clinicians' informational needs regarding the overall model objectives and attributes [18].

Moreover, our study explores the unique perspectives of GPs. In particular, GPs play a role as aggregators of a patient's medical information. As explored previously in [67] primary care consultation does not follow a linear structure and the needs of GPs and patients towards clinical documentation should evolve from this reality, which represents a challenge for automation. However, if implemented properly, these automation technologies can streamline GP's workflow, minimizing interruptions caused by interactions with the EHR system, and automating tasks, especially those falling out of the consultation's linear flow (e.g., having to stop the clinical interaction, to record something into the EHR). This would allow GPs to connect better with patients and decrease their cognitive load [27, 67].

5.1 Interpretation

From the identified themes and subthemes there are important lessons that can be related to previous research as well as new and emergent directions to explore in the future research.

5.1.1 Generally Positive View of Automation. Although clinicians raised concerns and challenges, they generally expressed a positive view of using text automation technologies in the general practice setting. When they were given specific examples, such as managing correspondence, transcribing a discharge letter, note-taking, or making appointments, clinicians perceived these low-value and time-consuming tasks to be more apt for automation. These findings are similar to previous reports that described that GPs found AI more appropriate for administrative tasks than for clinical reasoning tasks [9, 66]. System developers and designers should bear this in mind, as there is a potential for using these technologies in practice and the clinicians are less resistant than what could be expected.

5.1.2 Limited Concerns about a "Robot Takeover". A general perception among the interviewed GPs is that they were not particularly worried about machines overtaking general practice soon, as suggested in [124]. The GPs consider that they will still have an important role to play in the future, although perhaps different to the one they currently play. As AI, NLP, and the current wave of machine learning models mature, moving over the hype of superhuman AI into practical everyday tools addressing specific, sometimes tedious, and time-consuming tasks may prove to be a more useful framework for AI adoption.

5.1.3 AI literacy and Interface Design when Exposing AI. Our experience conducting the interviews is that clinicians valued the explanations and the ground-level introduction to AI and NLP. Previous studies have highlighted the importance of addressing AI literacy [73] and this is reflected by our approach of dedicating the initial part of our interviews to creating a shared understanding of the technology and its capabilities. Another important aspect is presenting how decision-making algorithms reach conclusions in a way that is transparent and understandable for users [24]. How these algorithms issue their recommendations and how this information is displayed to the user may have a strong effect on their success [58]. In our interviews, although there were no specific UI elements to display or interact with, we embedded AI actions into well-known clinical documentation tasks that helped clinicians imagine the specific role of AI in each task. This approach helped contextualize and focus clinicians on the value, challenges, and limitations that technologies with certain modes of user interaction may have. Researchers and designers should consider this when conducting studies and co-designing HCI elements for clinicians.

5.1.4 Reluctance to Give up Clinical Judgments and Uncertainty about Record-keeping Responsibilities. GPs envision themselves exercising a supervisory role over the automated outputs presented by AI. Clinicians consider that they will play a supervisory role and their judgement should prevail and overrule AI suggestions in case of disagreement. Previous literature has also suggested this supervisory role for future high-performance AI-assisted medicine [124]. Moreover, automating less cognitively demanding tasks, such as disease identification (information extraction) seems to be prioritized over establishing diagnoses or highly inferential tasks, such as suggesting a course of treatment. This finding corroborates previous findings on the intersection of GP and AI [9]. When designing NLP systems, developers must keep in mind that the end-user may want to maintain a high degree of control over the record and the user experience should be focused on facilitating these roles, without taking away their control.

5.1.5 Reconceptualize EHR Interface and Clinician-AI Interaction. The GP interviews showed a growing desire to modernize the EHR interface and interaction with health information and EHR software, which have been highlighted previously [4, 86]. Similarly, previous literature has also shown the importance of improving EHR visualization and its effect on clinicians' cognitive load [100]. To address this need, moving beyond the screen and keyboard, seems a crucial step, as GPs' day-to-day practice involves dealing with the computer screen and being detached from their patients. Previous studies have confirmed the growing time clinicians were "tethered to the EHR" [6] spending more time with the EHR than with the patients [95]. The GPs expressed general optimism that recent developments in HCI and newer interfaces and input-output devices may liberate them from the keyboard-and-screen paradigm in the future. Our findings also confirm an increasing HCI research interest in newer forms of interaction beyond the "Screen-Keyboard-Mouse" paradigm [72]. Several recent studies explored the potential of augmented reality and virtual reality applications in medicine [139]. In addition, the current episode-based EHR record, primarily inherited from paper-based records, may also be deemed obsolete, and moving beyond this type of record into a context-based [54] or semantically searchable record [30], could be a welcomed change.

5.1.6 Limits of AI and Practical Implementation Challenges. Besides the general optimism, certain concerns and challenges need to be addressed when NLP automation technologies are implemented in practice. Trust in AI in a high-stake, high-risk area such as medicine [32] poses one of the main barriers to implementation. How to increase clinicians' trust in AI systems, so that they are implemented safely into clinical practice remains an open question. Whether automated systems can be fully trusted and how this trust is earned and maintained is a key to their success. Equally, this is connected to the desire of keeping control of the record where an override option allows the clinician to regain control, edit, delete or modify the automated text output. In our work, clinicians envisage a model of progressive automation for overcoming the resistance to automated systems. This model can be applied to multiple levels of their interaction with technology. This gradual implementation with an initial testing and troubleshooting period could allow designing systems that are trusted by clinicians, as they evolve progressively from a fully supervised to a loosely supervised or even fully automated approach. This gradual implementation of technology has shown positive results previously as in the case of EHRs [113] where it helped overcome initial concerns that patients, clinicians and staff had regarding EHRs. Conversely, there have been numerous examples of poor implementation of technologies, and in particular, EHRs where it had a disruptive effect on clinical workflow, delaying care delivery [64] and increasing the time that clinicians spent in silence during the consultation with patients [42] as well as on decreasing satisfaction and increasing professional burnout [140]. Text automation solutions need to address the UX perspective not to create another click-fatigue scenario [28].

Another aspect of trust in AI is related to the use of AI technology on its own. It has been recently explored in NLP, where concerns about the widespread use and potential misuse of pre-trained language models have been raised [8, 84, 96]. Equally, the presence of biases in the training datasets of these language models is an ongoing issue in the NLP space [8]. Resolving the governance and addressing safety and privacy issues is only one dimension. An additional factor to be considered refers to the accountability of these algorithmic decisions. Although in the case of NLP algorithms and textual documentation, the decision-making step (prescribing, diagnosing, referring the patient) still refers to the clinician's judgement; the aspect of how the output of AI may influence the clinician and potentially cause errors and harm, and who is held accountable (clinician or algorithm), needs to be properly addressed before its implementation.

Additionally, an important factor to consider in terms of feasibility and implementation is how these newer AI-assisted systems may interact with the current records, and the willingness of practices, clinicians, and software companies to embrace change. This is particularly relevant in the case of proprietary record software and interoperability with legacy systems, especially if stakeholders do not perceive the need for improving them, so technological inertia [88] and market consolidation of the EHR industry [102] may hamper progress. If these issues are not properly addressed, clinicians may find a newly developed AI system hard to use in practice.

5.1.7 Explainable AI. Explainable AI has been gaining more traction and interest over recent years, especially on how a user can gain more understanding of how AI systems reached specific conclusions. Although there is a certain degree of limitation of black box AI systems (Adadi & Berrada, 2018), more recent approaches and developments shed some light on the inner workings of the algorithms (Adadi & Berrada, 2018). Although this is particularly relevant for systems that involve decisions (e.g., predictive models), it could be likewise applied to NLP systems and the context of clinical documentation.

5.1.8 The Challenge of Aggregation, Maintaining a Doctor's Sense of Control and data Accumulation. A more general consideration that emerged in our discussions with clinicians reflected on in the aggregation challenge's theme is the question about the need to record verbatim everything that happens in consultation. To what extent every conversation and free text needs to be recorded, transcribed, and quantified and if this has an overall positive effect on care is still an open debate. Having full verbatim records also paves the way for using AI as an auditing tool, where the system can retrospectively analyse and identify any misdiagnosis or overlooked information using NLP [66]. One emerging question is how to support doctors' sense of safety and control, so they would feel comfortable using these systems and not be disadvantaged by making errors or missing information. To allow this, responsible and ethical design considerations should be in place. Equally, this may be connected with the more general concerns regarding data accumulation itself, which may respond better to corporate interests, with their risks and dynamics [142], than be a pressing healthcare need. We believe, however, that a sensible implementation of text automation technologies focused on the clinicians' needs may still play a crucial role and more importantly, enable better patient care.

5.1.9 Solving Simpler Things Ahead of the Hard Questions. Applications of AI in medicine recently have focused on high-stake decisions, such as ruling out diseases and providing diagnosis [124] or question-answering systems and conversational agents [94]. However, clinicians also perceived a great value and potential healthcare improvements in solving simple or tedious tasks, which would not require innovative complex AI, but rather an application of existing technologies. Among these forms of automation could be voice recognition, text recognition engines, named entity extraction, and summarization. It could be argued that the recent developments in the NLP space paint this as rather an implementation and engineering challenge than an AI research problem that would require new AI and ML models.

5.1.10 Non-technical Solutions and the Need for Automation in Healthcare. We recognise that there are other non-automated options available such as the use of human medical scribes or introducing greater interoperability between systems and standards [83, 116]. However, we consider that these options have specific shortcomings and limitations which limit their applicability and do not overcome the need for text automation in EHRs.

Considering the use of human medical scribes, there are important limitations to their applicability including costs [111], frequent changes to the medical scribe workforce and a variable degree of satisfaction and availability [29]. More importantly, medical scribes are specific to the United States health sector, while many other healthcare systems make no use of them. Relying on an expensive, predominantly US-specific workaround may be of limited value for other healthcare systems, clinicians and EHR developers. That is potentially a reason why alternative strategies replacing or complementing human medical scribes have recently gained traction worldwide [27]. Moreover, in the specific setting of Primary Care, it is unrealistic to consider that a medical scribe workforce could be easily implemented to accompany every primary care, physician, and consultation while most healthcare systems struggle to retain their workforce [78].

Interestingly for UX researchers, the medical scribe industry perceives the improvement of EHR usability as a risk [29] and postulates an inverse relationship between EHR usability and the need for medical scribes. This could explain the perception that the use of medical scribes hampers the development of user-friendly and advanced EHR systems [41]. However, the existence of the medical scribe industry may also open opportunities to develop innovative technology solutions tailored specifically to this sector. For example, the work of Medical Scribes can be improved by the use of hybrid human/automated scribe solutions, where intelligent text automation technologies perform mundane tasks and streamline the work of their human counterparts. How to effectively combine the strengths of humans and machines in such a setting remains largely an under-explored area, with an opportunity for future UX research.

The text automation tasks related to medical correspondence (e.g., prescribing a drug suggested in a hospital discharge summary) could be overcome by better system integration and interoperability (e.g., a prescription from the hospital goes directly into primary care prescription record) bypassing the need for primary care clinicians doing a copyist work. However, several reasons limit the practical application of this simpler solution including cultural, structural and economic among others (Reisman, 2017). Although several initiatives are underway (Bonomi, 2016; Madhusoodanan, 2022), legal protection of health data still would limit the flow in various scenarios, making the need for systems that can extract and copy information from diverse sources, such as primary care, hospital records. More importantly, a higher degree of integration or even the use of a human workforce does not eliminate the need for coding and extracting clinical entities and relations; it just moves the problem to another part of the pipeline. Therefore, we consider that the use of a text-automation augmented EHR interface would be of value, regardless of the model of medical documentation in place.

5.1.11 Future of Clinical Consultation, Towards the Era of AI-assisted Clinical Practice. Besides all the concerns expressed, there is a positive attitude and anticipation that the clinical consultation may experience a general overhaul over the coming years. The experience of the years of the SARS-COV-2 pandemic with the spread of telehealth consultations (Monaghesh & Hajizadeh, 2020) and the growing flexibility in the options for consultation have paved the way for approaching further technological implementations into everyday practice. Equally, the general trend towards voice-assistant technologies, which have become ubiquitous over the past few years and the continued improvement of the NLP field fuelled by powerful multipurpose models (Brown et al., 2020), can help to materialize a near future where advanced AI systems take different roles in conjunction with the clinician. An all-powerful assistant that can take notes, modify prescriptions, create appointments, or refer patients to certain services, is within the reach of the current voice-assisted technologies, where for instance, creating a calendar appointment with a system such as Siri, has become a trivial task for any smartphone user. It is feasible to have such a system implemented successfully in the clinical setting, liberating resources, and reducing the administrative burden of the consultation.

The next step for AI systems and NLP is the creation of capable expert systems, which can make predictions and suggest to users the best course for managing a particular ailment or disease. The development of AI predictive models has exploded over recent years, although their quality and validation remain limited (Christodoulou et al., 2019). Translating NLP extracted entities into a pipeline that feeds predictive models, which are later converted into natural text is one of the possibilities that these technologies would allow. It remains unclear, however, how to make the best use of all the developed models, when only a handful of them are used in day-to-day medical practice. Moreover, how the clinician would react to having this external expert input and how the input would interact with their judgment and influence their decisions is still an open research area. Certainly, there is a growing need to explore these aspects from the prism of user experience, as they would offer the final layer between AI and clinical users and could either facilitate or trump the whole endeavour.

5.2 Limitations

This study is not without limitations. We have only included Australian GPs, although from various geographical areas, mostly surrounding the urban population centres. The needs of GPs practising in rural and remote areas may differ from those working in cities or big towns. Additionally, primary care and general practice settings may vary across regions and national/state health systems, pointing to potentially making the important features and the design options different for those GPs. Future developments such as user-centred design and prototyping need to include a broader representative sample of clinical professionals to further validate our findings.

Another potential limitation is that we did not collect GPs' prior familiarity with automation technologies and a certain degree of variability in terms of tech-savviness could have been present, especially taking into consideration that a broader sample could have different views on the use of EHR systems. To overcome this, we provided the GPs with a general explanatory introduction and a set of clinical documentation scenarios. Although we tried to minimise the interference between our views and those of the GPs, there could be a certain degree of influence in the responses of our participants due to the priming effect caused by these scenarios. We mitigated this by providing only general examples of well-known software platforms (e.g., Gmail's text auto-completion, Alexa voice assistant) reflecting what NLP is currently capable of and how these technologies can be deployed. Likewise, interviewing participants having no prior experience with the use of text automation technologies poses a limitation, as they may not completely understand the technology, its capabilities, or potential benefits. At the same time, we consider that this also offers advantages, as the participants are less primed to their previous experience and are the intended final users of these technologies. Another potential limitation that stems from this, is that although we based our presentation on the current capabilities of the technology in other areas, we presented the future NLP systems as if they have been tested and worked properly. When testing real-world systems and applications researchers may find different answers if the technology does not behave as expected or with sufficient accuracy.

Likewise, certain of the "consultation of the future" subthemes that emerged may appeal to a certain type of doctor, that is again, those more prone to technology use. GPs in different contexts or different practice styles may not find the appeal of having an AI as an expert system or as an assistant in their day-to-day practice.

Another potential limitation is that participants were recruited by expressing their interest in the topic and there is the risk that our sample might be biased toward more tech-savvy GPs. One of the potential challenges GPs mentioned is that older clinicians may rather exhibit a more hesitant attitude and increased resistance to change. Moreover, not only GPs work in primary care, but there are also nurses, allied health professionals, and administrative staff. Clinical and non-clinical users may have different needs and concerns surrounding the use of text automation and AI as well as sociocultural factors may influence their use [120]. EHR may not serve exclusively for clinical purposes, but depending on the health system, be used for billing purposes and secondary uses of healthcare data (such as epidemiological surveillance), thus requiring additional features and addressing different challenges. However, this should not impede the features that are important for doctors, the main users of the EHR. Future research will need to focus on addressing the needs of all potential users, and ideally, design a dynamic system that allows different modes of interaction, adapting to specific user needs.

Additionally, this study has focused on the primary care setting, whereas secondary care including hospitals and contexts like emergency medicine or intensive care units may require different features or focus on other types of interactions and benefits of AI and text automation. However, the advantages that can be obtained from information extraction and text summarization could be useful across the health system [48, 114] and could be used by a wide range of clinicians, specialities, and health settings.

Lastly, this study is a first step towards deriving the attitudes of clinicians and the required features of an NLP text automation system, to be designed and be practically useful. Further research is needed to validate these findings and address aspects ranging from the accuracy and validity of NLP algorithms to various aspects of user interaction and user experience. In addition, proper testing and evaluation of various modalities of interaction and the ways they instil trust and confidence in the system, without increasing the pervasive risks of automation bias, should be considered before any clinical validation study takes place. In our study, we have not focused on exploring other potential limitations and considerations surrounding the application of NLP to clinical texts. Additional work may be needed to explore the risks of algorithmic bias, especially concerning the use of large language models. This may include patient safety and legal issues surrounding the governance of confidential and sensitive medical information present in the EHR and decisions relying on unverified data automatically extracted from free-text sources.

6 CONCLUSIONS

We have identified four main themes surrounding the GPs' attitudes toward text automation. These four connect to the overarching idea of Doctor-AI collaboration and the synergistic relation between the doctor and automation technology. Thus, our work surfaces valuable insights, pertaining to the user-centred design of NLP tools for the medical domain, and primary care users more specifically.

Further research is needed to unravel how clinicians could interact with the automated text systems supported by a broader model of doctor-AI collaboration and how the new automation technologies could be designed and deployed safely and ethically in a real-world setting. Future research needs to focus on developing a deep understanding of doctor-AI collaboration models and designing medical NLP systems addressing the diverse needs and preferences of healthcare professionals while overcoming the errors and inefficiencies pervasive in the current EHR systems worldwide.

Future research needs to focus on exploring systems that perform specific NLP tasks and can be integrated into a cohesive EHR user experience. Developing modules that perform specific tasks such as named entity recognition or summarization and assessing how they interact with the clinical documentation task and integrate with the rest of the EHR system components is key to achieving a successful automation integration. Moreover, system designers need to evaluate different modes of interaction, such as those that allow different degrees of automation (from fully supervised to fully automated) to derive the ones most usable and best perceived by the user. Lastly, user experience needs to be designed around improving the ergonomics and facilitating seamless integration with the record, to avoid creating additional layers of complexity in the already cluttered interfaces and medical electronic systems.

Developers, researchers, and system designers alike should focus on addressing the pressing needs for text automation, especially in the more time-intensive areas and easier problems such as drug extraction rather than in the highly complex and intensive tasks such as clinical inference or diagnostics. Focusing on these low-hanging fruits may allow for further waves of progressive automation in the future. Moreover, there is an increasing need for a complete overhaul of the

EHR systems, addressing the need for changes to the interfaces and interactions that would allow meaningful and safe use of AI and text automation in clinical practice. System designers must address these needs when designing novel modes of interaction while bearing in mind that clinicians have to maintain a degree of control. Lastly, interactions between clinicians and the automated text technologies need to be designed with a focus on human-AI collaboration and not a substitution, as clinicians are here to stay for the near future, and it is only through this meaningful collaboration that we can ensure a beneficial change for the future of healthcare.

APPENDICES

APPENDIX 1

Initial technology pitch to participants:

Welcome and thanks for participating in this interview. Before we get started, I would like to give a short overview of the technologies we will be discussing today. As you may know, we have some new technologies that broadly we called “Artificial Intelligence” or AI that allow us to perform some complex operations and automate certain tasks. Examples of this are for instance self-driving cars, voice assistants such as Siri or Alexa, apps that can recognise faces in photos, such as in Facebook or Algorithms that can play games such as Chess or Go.

Among all these technologies, I would like to paint attention to some specific ones, to those AI algorithms that are focused on text and language. This are called “Natural Language Processing” algorithms or “NLP” for shorts. These algorithms. basic function is to recognise and process human, unprocessed text, such as what is present in spoken conversations, or in written pieces such as books, magazines, or letters. When narrowing it down to medicine, they have been used for a variety of purposes, for instance extracting relevant medical information (such as diagnoses or prescriptions from a text) to establish relations between different parts of a text such as a body part and a disease or for instance and in combination with some other clever technologies to convert voice into written text, and even produce some summary information of those texts. These are just a few examples, but generally speaking the idea behind NLP is to automatically “comprehend”, “organise” and even “produce” free text.

Now I would like to ask you some questions about your views on AI and implementing automation in General Practice and then in particular having in mind this text processing technologies.

This was then followed by the semi-structured interview questions.

APPENDIX 2

Box 1: General Questions Surrounding attitudes towards AI & automation

1. What do you think in general on the use of automation technologies (AI, NLP) in medicine?
2. What do you think specifically about the use of automation in your general practice?
3. What would be the most appropriate tasks (e.g., radiology, pathology, decision support system, triage, text analysis…) that could be done by automated systems?
4. What do you think are the main organizational barriers impeding the implementation of automation technologies in primary care?
5. What do you think could help simplify/smoothen its implementation in primary care?
6. Do you have any major concerns when using automated technologies in medicine? If yes, what are they?

Box 2: Specific questions regarding text automation and clinical documentation scenarios

8. How can automated clinical text processing be useful in your day-to-day practice? Why?
9. Are there any specific scenarios in your everyday use of electronic health records that you think automated clinical text processing would be particularly useful? Why?
10. Considering written medical information beyond your practice, such as clinical notes, discharge summaries, referral letters, what would be most useful tasks for these technologies?
11. In which situations such a system would be helpful (e.g., when first accessing patients notes, in complex patients, when reviewing and adding clinical letters to local EHRs…)? Why?
12. Is there particular information that you think those automated systems need to extract (e.g., medications, diseases, adverse drug reactions)? Why?
13. In which situations do you think such a system would potentially cause some problems or difficulties? Why?
14. Do you envision any other concerns with the use of automated text analysis technologies?
15. Could you think of other scenarios where this technology could be useful for you?
16. Can you try to describe how you imagine "the consultation of the future" in a few years, taking into consideration all the elements we discussed?

REFERENCES

[1] Amina Adadi and Mohammed Berrada. 2018. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access* 6, (2018), 52138–52160. https://ieeexplore.ieee.org/abstract/document/8466590.

[2] Emily Alsentzer, John Murphy, William Boag, Wei-Hung Weng, Di Jindi, Tristan Naumann, and Matthew McDermott. 2019. Publicly available clinical BERT Embeddings. In *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, Association for Computational Linguistics, Minneapolis, Minnesota, 72–78. DOI: https://doi.org/10.18653/v1/W19-1909

[3] Julia Amann, Alessandro Blasimme, Effy Vayena, Dietmar Frey, and Vince I. Madai. 2020. Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Med. Inform. Decis. Mak.* 20, 1 (2020), 1–9.

[4] Joseph Amlung, Hannah Huth, Theresa Cullen, and Thomas Sequist. 2020. Modernizing health information technology: Lessons from healthcare delivery systems. *JAMIA open* 3, 3 (2020), 369–377.

[5] Douglas E. Appelt. 1999. Introduction to information extraction. *Ai Commun.* 12, 3 (1999), 161–172.

[6] 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. *Ann. Fam. Med.* 15, 5 (2017), 419–426.

[7] Zachi I. Attia, Peter A. Noseworthy, Francisco Lopez-Jimenez, Samuel J. Asirvatham, Abhishek J. Deshmukh, Bernard J. Gersh, Rickey E. Carter, Xiaoxi Yao, Alejandro A. Rabinstein, and Brad J. Erickson. 2019. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet* 394, 10201 (2019), 861–867.

[8] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.* 610–623.

[9] Charlotte Blease, Ted J. Kaptchuk, Michael H. Bernstein, Kenneth D. Mandl, John D. Halamka, and Catherine M. DesRoches. 2019. Artificial intelligence and the future of primary care: Exploratory qualitative study of UK general practitioners' views. *J. Med. Internet Res.* 21, 3 (2019), e12802.

[10] Sabrina Bonomi. 2016. The electronic health record: a comparison of some European countries. In *Proceedings of the Information and Communication Technologies in Organizations and Society.* Springer, 33–50.

[11] Claus Bossen. 2006. Representations at work: a national standard for electronic health records. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work.* 69–78.

[12] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 2 (2006), 77–101.

[13] Erin Brodwin. 2021. Google is exploring a health record tool for patients. *STAT.* Retrieved November 22, 2022 from https://www.statnews.com/2021/04/09/google-health-records-app-care-studio/.

[14] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. (May 2020). Retrieved July 23, 2020 from https://arxiv.org/abs/2005.14165.

[15] Berry De Bruijn, Colin Cherry, Svetlana Kiritchenko, Joel Martin, and Xiaodan Zhu. 2011. Machine-learned solutions for three stages of clinical information extraction: The state of the art at i2b2 2010. *J. Am. Med. Informatics Assoc.* 18, 5 (2011), 557–562.

[16] Christoph Buck, Eileen Doctor, Jasmin Hennrich, Jan Jöhnk, and Torsten Eymann. 2022. General practitioners' attitudes toward artificial intelligence–enabled systems: Interview study. *J. Med. Internet Res.* 24, 1 (2022), e28916.

[17] Carrie J. Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S. Corrado, and Martin C. Stumpe. 2019. Human-centered tools for coping with imperfect algorithms during medical decision-making. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* 1–14.

[18] Carrie J. Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2019. " Hello AI": Uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. *Proc. ACM Human-computer Interact.* 3, CSCW (2019), 1–24.

[19] Åsa Cajander and Christiane Grünloh. 2019. Electronic health records are more than a work tool: Conflicting needs of direct and indirect stakeholders. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* 1–13.

[20] Kai-Wei Chang, Vinod Prabhakaran, and Vicente Ordonez. 2019. Bias and fairness in natural language processing. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing: Tutorial Abstracts.*

[21] Wendy W. Chapman, Prakash M. Nadkarni, Lynette Hirschman, Leonard W. D'avolio, Guergana K. Savova, and Ozlem Uzuner. 2011. Overcoming barriers to NLP for clinical text: The role of shared tasks and the need for additional creative solutions. *Journal of the American Medical Informatics Association* 18, 5 (2011), 540–543. DOI: https://doi.org/10.1136/amiajnl-2011-000465

[22] Basit Chaudhry, Jerome Wang, Shinyi Wu, Margaret Maglione, Walter Mojica, Elizabeth Roth, Sally C. Morton, and Paul G. Shekelle. 2006. Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. *Ann. Intern. Med.* 144, 10 (2006), 742–752.

[23] Jonathan H. Chen and Steven M. Asch. 2017. Machine learning and prediction in medicine—beyond the peak of inflated expectations. *N. Engl. J. Med.* 376, 26 (2017), 2507.

[24] Hao-Fei Cheng, Ruotong Wang, Zheng Zhang, Fiona O'Connell, Terrance Gray, F. Maxwell Harper, and Haiyi Zhu. 2019. Explaining decision-making algorithms through UI: Strategies to help non-expert stakeholders. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* 1–12.

[25] Munmun De Choudhury, Min Kyung Lee, Haiyi Zhu, and David A. Shamma. 2020. Introduction to this special issue on unifying human computer interaction and artificial intelligence. *Human–computer Interact.* 35, 5–6 (2020), 355–361.

[26] Enrico Coiera. 2015. *Guide to Health Informatics.* CRC Press.

[27] Enrico Coiera, Baki Kocaballi, John Halamka, and Liliana Laranjo. 2018. The digital scribe. *NPJ Digit. Med.* 1, 1 (2018), 1–5.

[28] Roger Collier. 2018. Rethinking EHR interfaces to reduce click fatigue and physician burnout.

[29] Sky Corby, Keaton Whittaker, Joan S. Ash, Vishnu Mohan, James Becton, Nicholas Solberg, Robby Bergstrom, Benjamin Orwoll, Christopher Hoekstra, and Jeffrey A. Gold. 2021. The future of medical scribes documenting in the electronic health record: results of an expert consensus conference. *BMC Med. Inform. Decis. Mak.* 21, 1 (2021), 1–11.

[30] Roberto Costumero, Consuelo Gonzalo, and Ernestina Menasalvas. 2014. TIDA: a spanish EHR semantic search engine. In *Proceedings of the 8th International Conference on Practical Applications of Computational Biology & Bioinformatics.* Springer, 235–242.

[31] Christine Dearnley. 2005. A reflection on the use of semi-structured interviews. *Nurse Res.* 13, 1 (2005).

[32] Matthew DeCamp and Jon C. Tilburt. 2019. Why we cannot trust artificial intelligence in medicine. *Lancet Digit. Heal.* 1, 8 (2019), e390.

[33] Dina Demner-Fushman, Wendy W. Chapman, and Clement J. McDonald. 2009. What can natural language processing do for clinical decision support? *J. Biomed. Inform.* 42, 5 (2009), 760–772.

[34] Li Deng. 2018. Artificial intelligence in the rising wave of deep learning: The historical path and future outlook [perspectives]. *IEEE Signal Process. Mag.* 35, 1 (2018), 177–180.

[35] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. DOI: https://doi.org/10.18653/v1/N19-1423
[36] George Edward Donovan. 1956. Automation in medicine. *Br. Med. J.* 2, 4983 (1956), 44.
[37] Chris Edwards. 2021. The best of NLP. *Commun. ACM* 64, 4 (2021), 9–11.
[38] Ceryn Evans and Jamie Lewis. 2018. *Analysing Semi-structured Interviews using Thematic Analysis: Exploring Voluntary Civic Participation Among Adults*. SAGE Publications Limited.
[39] Sarah Freymann Fontenot. 2013. The Affordable Care Act and electronic health care records. *Physician Exec.* 39, 6 (2013), 72–76.
[40] Lex Fridman, Li Ding, Benedikt Jenik, and Bryan Reimer. 2019. Arguing machines: Human supervision of black box AI systems that make life-critical decisions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.* 0.
[41] George A. Gellert, Ricardo Ramirez, and S. Luke Webster. 2015. The rise of the medical scribe industry: implications for the advancement of electronic health records. *Jama* 313, 13 (2015), 1315–1316.
[42] Darryl Gibbings-Isaac, Mudassar Iqbal, Mohammad Aumran Tahir, Pushpa Kumarapeli, and Simon de Lusignan. 2012. The pattern of silent time in the clinical consultation: an observational multichannel video study. *Fam. Pract.* 29, 5 (2012), 616–621.
[43] F. J. Gilbert, S. W. Smye, and C. B. Schönlieb. 2020. Artificial intelligence in clinical imaging: a health system approach. *Clin. Radiol.* 75, 1 (2020), 3–6.
[44] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press.
[45] W. Gossman, V. Lew, and S. Ghassemzadeh. 2020. SOAP Notes. *StatPearls [Internet].* https://www.ncbi.nlm.nih.gov/books/NBK482263/.
[46] Greg Guest, Arwen Bunce, and Laura Johnson. 2006. How many interviews are enough? An experiment with data saturation and variability. *Field Methods* 18, 1 (2006), 59–82.
[47] Greg Guest, Emily Namey, and Mario Chen. 2020. A simple method to assess and report thematic saturation in qualitative research. *PLoS One* 15, 5 (2020), e0232076.
[48] Udo Hahn and Michel Oleynik. 2020. Medical information extraction in the age of deep learning. *Yearb. Med. Inform.* 29, 1 (2020), 208–220. DOI: https://doi.org/10.1055/S-0040-1702001
[49] Kai Hakala and Sampo Pyysalo. 2019. Biomedical named entity recognition with multilingual BERT. In *Proceedings of the 5th Workshop on BioNLP Open Shared Tasks.* 56–61.
[50] Bruce Hanington and Bella Martin. 2019. *Universal Methods of Design Expanded and Revised: 125 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*. Rockport Publishers.
[51] Yun He, Ziwei Zhu, Yin Zhang, Qin Chen, and James Caverlee. 2020. Infusing disease knowledge into BERT for health question answering, medical inference and disease name recognition. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Association for Computational Linguistics, Online, 4604–4614. DOI: https://doi.org/10.18653/v1/2020.emnlp-main.372
[52] Kasper Hornbæk and Erik Frøkjær. 2003. Reading patterns and usability in visualizations of electronic documents. *ACM Trans. Comput. Interact.* 10, 2 (2003), 119–149.
[53] Willem Horst and Ben Matthews. 2016. Live prototyping. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems.* 632–642.
[54] William Hsu, Ricky K. Taira, Suzie El-Saden, Hooshang Kangarloo, and Alex A. T. Bui. 2012. Context-based electronic health record: toward patient specific healthcare. *IEEE Trans. Inf. Technol. Biomed.* 16, 2 (2012), 228–234.
[55] Edwin Hutchins, Barbara Holder, and Michael Hayward. 1999. Pilot attitudes towards automation. Retrieved November 22, 2022 from https://pages.ucsd.edu~ehutchins/aviation/attitudes/attitudes.pdf.
[56] Stanislav Ivanov, Mihail Kuyumdzhiev, and Craig Webster. 2020. Automation fears: drivers and solutions. *Technol. Soc.* 63, (2020), 101431.
[57] Maia Jacobs, Jeffrey He, Melanie F. Pradier, Barbara Lam, Andrew C. Ahn, Thomas H. McCoy, Roy H. Perlis, Finale Doshi-Velez, and Krzysztof Z. Gajos. 2021. Designing AI for trust and collaboration in time-constrained medical decisions: A sociotechnical lens. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–14.
[58] Maia Jacobs, Melanie F. Pradier, Thomas H. McCoy, Roy H. Perlis, Finale Doshi-Velez, and Krzysztof Z. Gajos. 2021. How machine-learning recommendations influence clinician treatment selections: the example of antidepressant selection. *Transl. Psychiatry* 11, 1 (2021), 1–9.
[59] E. W. Jamoom, D. Heisey-Grove, N. Yang, and P. Scanlon. 2016. Physician opinions about EHR use by EHR experience and by whether the practice had optimized its EHR use. *J. Heal. Med. informatics* 7, 4 (2016).

[60] Christian P. Janssen, Stella F. Donker, Duncan P. Brumby, and Andrew L. Kun. 2019. History and future of human-automation interaction. *Int. J. Hum. Comput. Stud.* 131, (2019), 99–107.

[61] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, and Anna Potapenko. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 7873 (2021), 583–589.

[62] Florian Jungmann, Tobias Jorg, Felix Hahn, Daniel Pinto Dos Santos, Stefanie Maria Jungmann, Christoph Düber, Peter Mildenberger, and Roman Kloeckner. 2021. Attitudes toward artificial intelligence among radiologists, IT specialists, and industry. *Acad. Radiol.* 28, 6 (2021), 834–840.

[63] Danai Khemasuwan, Jeffrey S. Sorensen, and Henri G. Colt. 2020. Artificial intelligence in pulmonary medicine: Computer vision, predictive model and COVID-19. *Eur. Respir. Rev.* 29, 157 (2020).

[64] Mi Ok Kim, Enrico Coiera, and Farah Magrabi. 2017. Problems with health information technology and their effects on care delivery and patient outcomes: A systematic review. *J. Am. Med. Informatics Assoc.* 24, 2 (2017), 246–250.

[65] Will Knight. 2022. Microsoft Makes a $16 Billion Entry Into Health Care AI. *Wired* (May 2022). Retrieved November 22, 2022 from https://www.wired.com/story/microsoft-dollar16-billion-entry-health-care-ai/.

[66] A. Baki Kocaballi, Kiran Ijaz, Liliana Laranjo, Juan C. Quiroz, Dana Rezazadegan, Huong Ly Tong, Simon Willcock, Shlomo Berkovsky, and Enrico Coiera. 2020. Envisioning an artificial intelligence documentation assistant for future primary care consultations: A co-design study with general practitioners. *J. Am. Med. Informatics Assoc.* 27, 11 (2020), 1695–1704.

[67] Ahmet Baki Kocaballi, Enrico Coiera, Huong Ly Tong, Sarah J. White, Juan C. Quiroz, Fahimeh Rezazadegan, Simon Willcock, and Liliana Laranjo. 2019. A network model of activities in primary care consultations. *J. Am. Med. Informatics Assoc.* 26, 10 (2019), 1074–1082.

[68] Anssi Lensu. 2002. Computationally intelligent methods for qualitative data analysis. *Jyväskylä Studies in Computing* 23 (2002). Retrieved November 22, 2022 from https://jyx.jyu.fi/handle/123456789/13284.

[69] Brenna Li, Noah Crampton, Thomas Yeates, Yu Xia, Xirong Tian, and Khai Truong. 2021. automating clinical documentation with digital scribes: Understanding the impact on physicians. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–12.

[70] Steven Y. Lin, Megan R. Mahoney, and Christine A. Sinsky. 2019. Ten ways artificial intelligence will transform primary care. *J. Gen. Intern. Med.* 34, 8 (2019), 1626–1630.

[71] Steven Y. Lin, Tait D. Shanafelt, and Steven M. Asch. 2018. Reimagining clinical documentation with artificial intelligence. In *Proceedings of the Mayo Clinic Proceedings.* Mayo Foundation for Medical Education and Research, 563–565.

[72] Gitte Lindgaard and Sheila Narasimhan. 2009. Mobile HCI: thinking beyond the screen-keyboard-mouse interaction paradigm. *Int. J. Mob. Hum. Comput. Interact.* 1, 3 (2009), 46–60.

[73] Duri Long and Brian Magerko. 2020. What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–16.

[74] M. De Luca, C. S. Szostek, J. Cartlidge, and Dave Cliff. 2011. *Studies of Interaction Between Human Traders and Algorithmic Trading Systems.* UK Government Office for Science.

[75] David Lyell and Enrico Coiera. 2017. Automation bias and verification complexity: a systematic review. *J. Am. Med. Informatics Assoc.* 24, 2 (2017), 423–431.

[76] David Lyell, Farah Magrabi, Magdalena Z. Raban, Lisa G. Pont, Melissa T. Baysari, Richard O. Day, and Enrico Coiera. 2017. Automation bias in electronic prescribing. *BMC Med. Inform. Decis. Mak.* 17, 1 (2017), 1–10.

[77] Karl F. MacDorman, Sandosh K. Vasudevan, and Chin-Chang Ho. 2009. Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *AI Soc.* 23, 4 (2009), 485–510.

[78] Logan MacLean, Susan Hassmiller, Franklin Shaffer, Kathleen Rohrbaugh, Tiffany Collier, and Julie Fairman. 2014. Scale, causes, and implications of the primary care nursing shortage. *Annu. Rev. Public Health* 35, (2014), 443–457.

[79] Jyoti Madhusoodanan. 2022. Unlocking the potential of health data to help research and treatments. *Nature* 605, 7908 (2022), 182–183.

[80] Bill Manaris. 1998. Natural language processing: A human-computer interaction perspective. In *Proceedings of the Advances in Computers.* Elsevier, 1–66.

[81] Inderjeet Mani and Mark T. Maybury. 2001. Automatic summarization. *Advances in Automatic Text Summarization*, Advaith Siddhartan. Inderjeet Mani and Mark T. Maybury (Eds.). MIT Press, pp. $47.95/£32.95 (paperback). Natural Language Engineering 7, 3 (2001), 271–274. DOI: https://doi.org/10.1017/S1351324901222700

[82] Jörg Hans Mayer, Oliver Stritzel, Markus Esswein, and Reiner Quick. 2020. Towards Natural Language Processing: An Accounting Case Study. In *Proceedings of the ICIS.*

[83] Clement J. McDonald. 1997. The barriers to electronic medical record systems and how to overcome them. *J. Am. Med. Informatics Assoc.* 4, 3 (1997), 213–221.

[84] Kris McGuffie and Alex Newhouse. 2020. The radicalization risks of GPT-3 and advanced neural language models. https://www.middlebury.edu/institute/academics/centers-initiatives/ctec/ctec-publications/radicalization-risks-gpt-3-and-neural-language.

[85] Edward R. Melnick, Liselotte N. Dyrbye, Christine A. Sinsky, Mickey Trockel, Colin P. West, Laurence Nedelec, Michael A. Tutty, and Tait Shanafelt. 2020. The association between perceived electronic health record usability and professional burnout among US physicians. In *Proceedings of the Mayo Clinic Proceedings*, Elsevier, 476–487.

[86] Blackford Middleton, Meryl Bloomrosen, Mark A. Dente, Bill Hashmat, Ross Koppel, J. Marc Overhage, Thomas H. Payne, S. Trent Rosenbloom, Charlotte Weaver, and Jiajie Zhang. 2013. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. *J. Am. Med. Informatics Assoc.* 20, e1 (2013), e2–e8.

[87] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space BT-1st international conference on learning representations. In *International Conference on Learning Representations (ICLR'13, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings).* Retrieved from http://arxiv.org/abs/1301.3781.

[88] Joel Mokyr. 1992. Technological inertia in economic history. *J. Econ. Hist.* 52, 2 (1992), 325–338.

[89] Luiz Moutinho and Anne Smith. 2000. Modelling bank customer satisfaction through mediation of attitudes towards human and automated banking. *Int. J. bank Mark.* 18, 3 (2000), 124–134. DOI: https://doi.org/10.1108/02652320010339699

[90] Alison R. Murphy, Madhu C. Reddy, and Nathan J. McNeese. 2014. Exploring the perceptions and use of electronic medical record systems by non-clinicians. In *Proceedings of the 2014 Conference on Designing Interactive Systems.* 429–432.

[91] David Nadeau and Satoshi Sekine. 2007. A survey of named entity recognition and classification. *Lingvisticae Investig.* 30, 1 (2007), 3–26.

[92] Constanza L. Andaur Navarro, Johanna A. A. Damen, Toshihiko Takada, Steven W. J. Nijman, Paula Dhiman, Jie Ma, Gary S. Collins, Ram Bajpai, Richard D. Riley, and Karel G. M. Moons. 2021. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *bmj* 375, (2021), n2281. https://www.bmj.com/content/375/bmj.n2281.

[93] Lucila Ohno-Machado. 2011. Realizing the full potential of electronic health records: The role of natural language processing. *J. Am. Med. Informatics Assoc.* 18, 5 (2011), 539.

[94] David Oliver. 2019. David Oliver: Lessons from the Babylon Health saga. *BMJ* 365, (2019). https://www.bmj.com/content/365/bmj.l2387.

[95] J. Marc Overhage and David McCallie Jr. 2020. Physician time spent using the electronic health record during outpatient encounters: A descriptive study. *Ann. Intern. Med.* 172, 3 (2020), 169–174.

[96] Xudong Pan, Mi Zhang, Shouling Ji, and Min Yang. 2020. Privacy risks of general-purpose language models. In *Proceedings of the 2020 IEEE Symposium on Security and Privacy.* IEEE, 1314–1331.

[97] Rushikesh Pandya, Shrey Nadiadwala, Rajvi Shah, and Manan Shah. 2020. Buildout of methodology for meticulous diagnosis of K-complex in EEG for aiding the detection of Alzheimer's by artificial intelligence. *Augment. Hum. Res.* 5, 1 (2020), 1–8.

[98] Claus Bossen and Kathleen H. Pine. 2022. Batman and robin in healthcare knowledge work: Human-AI collaboration by clinical documentation integrity specialists. *ACM Trans. Comput.-Hum. Interact.* (2022). DOI: https://doi.org/10.1145/3569892

[99] S. Polesie, M. Gillstedt, H. Kittler, A. Lallas, P. Tschandl, I. Zalaudek, and J. Paoli. 2020. Attitudes towards artificial intelligence within dermatology: An international online survey. *British Journal of Dermatology* 183, 1 (2020), 159–161. DOI: https://doi.org/10.1111/bjd.18875

[100] Ari H. Pollack and Wanda Pratt. 2020. Association of health record visualizations with physicians' cognitive load when prioritizing hospitalized patients. *JAMA Netw. open* 3, 1 (2020), e1919301–e1919301.

[101] Fred Popowich. 2005. Using text mining and natural language processing for health care claims processing. *ACM SIGKDD Explor. Newsl.* 7, 1 (2005), 59–66.

[102] Mary K. Pratt. 2019. EHR market consolidation and the impact on physicians. *Med. Econ.* 96, 1 (2019). Retrieved October 27, 2021 from https://www.medicaleconomics.com/view/ehr-market-consolidation-and-impact-physicians.

[103] Affordable Care Act. 2010. Patient protection and affordable care act. *Public Law* 111 (2010), 148.

[104] James Proudman. 2018. Cyborg supervision–the application of advanced analytics in prudential supervision. In *Proceedings of the Speech, Workshop on Research on Bank Supervision,* London.

[105] Juan C. Quiroz, Liliana Laranjo, Ahmet Baki Kocaballi, Shlomo Berkovsky, Dana Rezazadegan, and Enrico Coiera. 2019. Challenges of developing a digital scribe to reduce clinical documentation burden. *NPJ Digit. Med.* 2, 1 (2019), 1–6.

[106] Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, Rachel Prudden, Amol Mandhane, Aidan Clark, Andrew Brock, Karen Simonyan, Raia Hadsell, Niall Robinson, Ellen Clancy, Alberto Arribas, and Shakir Mohamed. 2021. Skilful precipitation nowcasting using deep generative models of radar. *Nature* 597, 7878 (2021), 672–677. DOI: https://doi.org/10.1038/s41586-021-03854-z

[107] Miriam Reisman. 2017. EHRs: The challenge of making electronic data usable and interoperable. *Pharm. Ther.* 42, 9 (2017), 572.

[108] Benjamin J. Roberts, Steven L. Gordon, Jarè Struwig, Narnia Bohler-Muller, and Michael Gastrow. 2022. Promise or precarity? South African attitudes towards the automation revolution. *Dev. South. Afr.* 39, 4 (2022), 498–515. DOI: https://doi.org/10.1080/0376835X.2021.1978932

[109] Kate Roberts, Anthony Dowell, and Jing-Bao Nie. 2019. Attempting rigour and replicability in thematic analysis of qualitative research data; a case study of codebook development. *BMC Med. Res. Methodol.* 19, 1 (2019), 66. DOI: https://doi.org/10.1186/s12874-019-0707-y

[110] Abeed Sarker, Yuan-Chi Yang, Mohammed Ali Al-Garadi, and Aamir Abbas. 2020. A light-weight text summarization system for fast access to medical evidence. *Front. Digit. Heal.* 2, (2020). https://www.frontiersin.org/articles/10.3389/fdgth.2020.585559/full.

[111] Gordon D. Schiff and Laura Zucker. 2016. Medical scribes: salvation for primary care or workaround for poor EMR usability? *Journal of General Internal Medicine* 31, 9 (2016), 979–981.

[112] Tina Shah, Andrea Borondy Kitts, Jeffrey A. Gold, Keith Horvath, Alex Ommaya, Frank Opelka, Luke Sato, Gretchen Schwarze, Mark Upton, and Lew Sandy. 2020. Electronic Health Record Optimization and Clinician Well-Being: A Potential Roadmap Toward Action. *NAM Perspect.* (2020). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8916811/.

[113] Renée R. Shield, Roberta E. Goldman, David A. Anthony, Nina Wang, Richard J. Doyle, and Jeffrey Borkan. 2010. Gradual electronic health record implementation: New insights on physician and patient adaptation. *Ann. Fam. Med.* 8, 4 (2010), 316–326.

[114] Yuqi Si and Kirk Roberts. 2018. A frame-based NLP system for cancer-related information extraction. In *Proceedings of the AMIA Annual Symposium Proceedings.* American Medical Informatics Association, 1524.

[115] Dean F. Sittig. 2013. *Electronic Health Records: Challenges in Design and Implementation.* CRC Press.

[116] Dean F. Sittig and Adam Wright. 2015. What makes an EHR "open" or interoperable? *J. Am. Med. Informatics Assoc.* 22, 5 (2015), 1099–1101.

[117] Ewout W. Steyerberg. 2019. *Clinical Prediction Models.* Springer.

[118] Reed T. Sutton, David Pincock, Daniel C. Baumgart, Daniel C. Sadowski, Richard N. Fedorak, and Karen I. Kroeker. 2020. An overview of clinical decision support systems: Benefits, risks, and strategies for success. *NPJ Digit. Med.* 3, 1 (2020), 1–10.

[119] Tania Tajirian, Vicky Stergiopoulos, Gillian Strudwick, Lydia Sequeira, Marcos Sanches, Jessica Kemp, Karishini Ramamoorthi, Timothy Zhang, and Damian Jankowicz. 2020. The influence of electronic health record use on physician burnout: Cross-sectional survey. *J. Med. Internet Res.* 22, 7 (2020), e19274.

[120] Ayushi Tandon. 2019. Variations in the use of electronic medical records: Role of sociocultural aspects. In *Proceedings of the Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing.* 94–97.

[121] Gareth Terry, Nikki Hayfield, Victoria Clarke, and Virginia Braun. 2017. Thematic analysis. *SAGE Handb. Qual. Res. Psychol.* 2, (2017), 17–37. https://uwerepository.worktribe.com/output/888518/thematic-analysis.

[122] Anja Thieme, Danielle Belgrave, and Gavin Doherty. 2020. Machine learning in mental health: A systematic review of the HCI literature to support the development of effective and implementable ML systems. *ACM Trans. Comput. Interact.* 27, 5 (2020), 1–53.

[123] Eric Topol. 2019. *Deep Medicine: How Artificial Intelligence can Make Healthcare Human Again.* Hachette UK.

[124] Eric J. Topol. 2019. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* 25, 1 (2019), 44–56. DOI: https://doi.org/10.1038/s41591-018-0300-7

[125] Ozlem Uzuner, Brett R. South, Shuying Shen, and Scott L. DuVall. 2011. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *J. Am. Med. Inform. Assoc.* 18, 5 (2011), 552–556. DOI: https://doi.org/10.1136/amiajnl-2011-000203

[126] Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, and Harshit Surana. 2020. *Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems.* O'Reilly Media.

[127] Rafael Valencia-García and Francisco García-Sánchez. 2013. NATURAL LANGUAGE PROCESSING and Human–Computer Interaction. *Comput. Stand. Interfaces* 5, 35 (2013), 415–416.

[128] Neeraj Varshney, Swaroop Mishra, and Chitta Baral. 2020. It's better to say "I can't answer" than answering incorrectly: Towards safety critical NLP systems. arXiv:2008.09371. Retrieved from https://arxiv.org/abs/2008.09371.

[129] Effy Vayena, Alessandro Blasimme, and I. Glenn Cohen. 2018. Machine learning in medicine: addressing ethical challenges. *PLoS Med.* 15, 11 (2018), e1002689. https://journals.plos.org/plosmedicine/article/citation?id=10.1371/journal.pmed.1002689.

[130] Robert Wachter and Jeff Goldsmith. 2018. To combat physician burnout and improve care, fix the electronic health record. *Harv. Bus. Rev.* (2018). https://hbr.org/2018/03/to-combat-physician-burnout-and-improve-care-fix-the-electronic-health-record.

[131] Howard Waitzkin. 1989. A critical theory of medical discourse: Ideology, social control, and the processing of social context in medical encounters. *J. Health Soc. Behav.* 30, 2 (1989), 220–239.

[132] Dakuo Wang, Elizabeth Churchill, Pattie Maes, Xiangmin Fan, Ben Shneiderman, Yuanchun Shi, and Qianying Wang. 2020. From human-human collaboration to human-ai collaboration: Designing ai systems that can work together with people. In *Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–6.

[133] Dakuo Wang, Liuping Wang, Zhan Zhang, Ding Wang, Haiyi Zhu, Yvonne Gao, Xiangmin Fan, and Feng Tian. 2021. "Brilliant AI Doctor" in Rural Clinics: Challenges in AI-Powered Clinical Decision Support System Deployment. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–18.

[134] Quentin Waymel, Sammy Badr, Xavier Demondion, Anne Cotten, and Thibaut Jacques. 2019. Impact of the rise of artificial intelligence in radiology: What do radiologists think? *Diagn. Interv. Imaging* 100, 6 (2019), 327–336.

[135] Lawrence L. Weed. 1969. *Medical Records, Medical Education, and Patient Care: The Problem-oriented Record as a Basic Tool.* Press of Case Western Reserve University.

[136] Astrid Weiss, Regina Bernhaupt, Daniel Dürnberger, M. Altmaninger, Roland Buchner, and M. Tscheligi. 2010. User Experience Evaluation with a Wizard of Oz approach: Technical and Methodological Considerations. DOI:https://doi.org/10.1109/ICHR.2009.5379559

[137] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: state-of-the-art natural nanguage nrocessing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Association for Computational Linguistics, Online, 38–45. DOI:https://doi.org/10.18653/v1/2020.emnlp-demos.6

[138] Stephen Wu, Kirk Roberts, Surabhi Datta, Jingcheng Du, Zongcheng Ji, Yuqi Si, Sarvesh Soni, Qiong Wang, Qiang Wei, and Yang Xiang. 2020. Deep learning in clinical natural language processing: a methodical review. *J. Am. Med. Informatics Assoc.* 27, 3 (2020), 457–470.

[139] Andy Wai Kan Yeung, Anela Tosevska, Elisabeth Klager, Fabian Eibensteiner, Daniel Laxar, Jivko Stoyanov, Marija Glisic, Sebastian Zeiner, Stefan Tino Kulnik, and Rik Crutzen. 2021. Virtual and augmented reality applications in medicine: analysis of the scientific literature. *J. Med. internet Res.* 23, 2 (2021), e25499.

[140] Kai Zheng, Raj M. Ratwani, and Julia Adler-Milstein. 2020. Studying workflow and workarounds in electronic health record–supported work to improve health system performance. *Ann. Intern. Med.* 172, 11_Supplement (2020), S116–S122.

[141] Xiaomu Zhou, Kai Zheng, Mark Ackerman, and David Hanauer. 2012. Cooperative documentation: the patient problem list as a nexus in electronic health records. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work.* 911–920.

[142] Shoshana Zuboff. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* Profile Books.

[143] 2022. MedicalDirector - Software Solutions for Medical Practitioners. *MedicalDirector.* Retrieved November 22, 2022 from https://www.medicaldirector.com.

Received 12 November 2021; revised 27 May 2022; accepted 29 August 2022