Intelligent Human Systems Integration (IHSI 2025), Vol. 160, 2025, 297–306
https://doi.org/10.54941/ahfe1005835

The Lack of Explainability in Automatic Speech Recognition Can Cause Faux Data Work

Silja Vase

Department of Communication, University of Copenhagen, Copenhagen, Denmark

ABSTRACT

Automatic Speech Recognition (ASR) is increasingly used in healthcare to reduce documentation workloads by transcribing spoken words into Electronic Health Records (EHRs). However, these systems, based on machine learning, require ongoing data annotation and validation by healthcare professionals to ensure accuracy. This paper, based on fieldwork at a public Danish hospital, investigates the challenges healthcare professionals face in detecting and addressing technical issues, such as glitches, within ASR systems. Using mixed methods, the study reveals that healthcare professionals spend significant time annotating and training the machine learning algorithms—time that could otherwise be dedicated to patient care. Without access to clear metrics, like recognition rates, healthcare professionals are unable to effectively evaluate their data annotating efforts, leading to "faux data work," where data tasks seem productive but fail to improve system performance. The paper proposes two strategies to mitigate this issue; 1) providing transparent system metrics to enhance user engagement; and 2) creating structured sites of collaboration between healthcare professionals and IT professionals for better reporting of technical issues. These solutions aim to reduce inefficiencies and improve ASR accuracy in clinical settings.

Keywords: Automatic speech recognition, Faux data work, Transparency, Glitches

INTRODUCTION

Amid increasing efforts to alleviate workloads and streamline documentation processes in healthcare, Automatic Speech Recognition (ASR) has become a prominent tool in hospitals (Adedeji *et al.*, 2024). ASR technology transcribes spoken words into written text for integration into Electronic Healthcare Records (EHRs), assisting healthcare professionals (HCPs), such as physicians, in enhancing the efficiency and speed of clinical documentation (Kumar, 2024). Recent advancements in ASR technology have focused on improving recognition accuracy, with English ASR models demonstrating recognition rates exceeding 95% (Aldarmaki *et al.*, 2022; Baevski *et al.*, 2022; Chen *et al.*, 2022), aligning with the commercial minimum acceptance threshold (Lewis, 2016). This level of accuracy corresponds to fewer than 5 transcription errors per 100 words. However, real-world applications often reveal higher error rates, particularly in non-English contexts or when assisting diverse linguistic groups (Ngueajio and Washington, 2022; Terriza *et al.*, 2022; Vase and Berget, 2023). Maintaining high recognition accuracy requires ongoing machine learning (ML) training as the system encounters novel inputs during operational use (Hacking *et al.*, 2023; Kumar, 2024). Critical to this training process is the ability of users to manage technical challenges, including system glitches and bugs (Lee and Kolodge, 2020; Kaswan *et al.*, 2021). A growing body of research underscores the need for real-world usability studies to address these issues, particularly in the context of ASR systems deployed in clinical environments (Palivela *et al.*, 2023; Turri and Dzombak, 2023). These emphasize the importance of incorporating diverse linguistic data into language models that underpin ML processes (Meripo and Konam, 2022; Hacking *et al.*, 2023). They also highlight the necessity for organizations to empower users to identify and resolve system failures autonomously, ensuring sustained system reliability (Lin and Jackson, 2023). Despite the recognized importance of diversity in ASR systems—particularly in public domains where they must accommodate varied user voices (Vase and Berget, 2023)—this remains a largely underexplored area (Palivela *et al.*, 2023). Given this gab, and the inevitability of operational issues of such complex technologies in practice (Raji *et al.*, 2022) this research asks: *How do healthcare professionals detect technical issues in ASR systems, and what impact do these issues have on reducing workload of documentation workflows?* This paper aims to address this research question by mixed methods examining the use of ASR in a Danish public hospital where it has been utilized for over a decade (Alapetite *et al.*, 2009; Vase, 2021).

BACKGROUND

ASR is increasingly being implemented in hospitals worldwide, promising faster and more seamless documentation (Adedeji *et al.,* 2024). ASR systems rely on ML techniques to convert spoken language into text by predicting sequences of words based on prior context ML (Lewis, 2016; Pascual *et al.,* 2024). Central to these systems is the language model, which serves as a foundational database of linguistic knowledge, determining the probability of word combinations based on training data (Lewis, 2016; Meripo and Konam, 2022). For example, when a physician says, "the patient exhibits symptoms of...," the model predicts subsequent words like "fatigue" or "infection" based on its prior training. While ASR systems perform well in general applications, their efficacy in specialized domains such as healthcare is often limited (Ngueajio and Washington, 2022; Terriza *et al.*, 2022; Hacking *et al.*, 2023). Language models are typically pre-trained on generic datasets, which fail to capture the specialized vocabulary, evolving terminology, and diverse accents prevalent in clinical environments (Adedeji *et al.,* 2024; Pascual *et al.,* 2024). Consequently, HCPs frequently encounter transcription errors, particularly when using everyday terms like "plane" or "plant" or speaking with dialects or pitches underrepresented in the training data. In documenting EHRs, HCPs annotate data when correcting transcription errors and providing feedback to improve the ML algorithms underlying ASR systems (Adedeji *et al.,* 2024). This process enables developers to refine language models and enhance their inclusivity. However, it also reflects an increasing reliance on HCPs to perform "data work," a term denoting the sociotechnical tasks that bridge human expertise and computational systems (Bossen *et al.*, 2019). This shift expands the scope of HCPs' responsibilities, incorporating technical oversight into their roles (Bossen and Bertelsen, 2023; Bertelsen *et al.*, 2024). Yet, research on ASR users identifying glitches is limited and discussed by non-formal roles such as "transcription reviewers" encircling diverse professionals examining potential ASR inaccuracies (Adedeji *et al.*, 2024). This limitation is particularly pronounced in smaller language settings (Palivela *et al.*, 2023; Vase and Berget, 2023), such as Denmark.

Hard-to-Explain Technical Issues

ML systems often receive extensive attention during the initial phases of development, particularly in addressing bugs and glitches. However, once these systems are deployed in real-world environments, this level of scrutiny diminishes significantly, leading to risks of poor data practices (Raji *et al.*, 2022). The real-world deployment phase introduces unforeseen challenges as systems interact with diverse and complex datasets, often resulting in glitches and other technical issues (Lin and Jackson, 2023). These are not always evident during controlled testing and can lead to misclassifications and inaccurate outputs (Herrmann and Pfeiffer, 2023). This issue is exacerbated by the expectation that users interact effectively with these technologies, often without adequate training or institutional support (Yao *et al.*, 2024). To address these challenges, ongoing refinement of ML systems is necessary, drawing heavily on the perspectives and experiences of end-users who interact with the system in practice (Haque and Rubya, 2022; Bi and Huang, 2023). User feedback becomes critical in identifying limitations and providing actionable insights to adapt and sustain these systems effectively. Rather than being static and objective, data is understood as a dynamic entity, constantly undergoing reconstitution, repair, and adaptation through interactions with users and systems. To this extent, Lin and Jackson (2023) point to "sites of collaboration" referring to locations or points in the workflow where different actors come together to address, negotiate, and resolve errors. This perspective shifts the focus from a reliance on static quantitative metrics to a more nuanced understanding of how data evolves in real-world settings—a limited approach in literature on ASR in healthcare (Hacking *et al.*, 2023).

METHOD AND CASE

HCPs verbally dictate notes, which the ASR transcribes into text in real time using probabilistic calculations. HCPs then correct any mistranslations by labeling the errors, training the ASR system to improve future recognition. The labeled data is forwarded to medical secretaries for validation, who then send it to external suppliers to update the language model used by the ASR.

This study employs a mixed-methods approach to analyze HCPs utilizing ASR. Confidence intervals were used to provide a range of values (Lott and Reiter, 2020) to illustrate recognition rates. These rates measure mistranslated words or sentences when an HCP corrects these and indicate the accuracy of the ASR in transcribing spoken language. This assessment allows to determine whether the ASR achieves the commercial minimum acceptance rate of 95% (Lewis, 2016). The Wilson confidence interval for the *p*-proportion in a binomial distribution, offers advantages for repeated sampling properties (Lott & Reiter, 2020). Intervals are illustrated by the proportion of the mistranslated words per medical record calculated by the confidence intervals for the probability parameter *p* as illustrated in Equation 1. A sample of more than 70 notes per HCP (n = 220) was observed and collected to create a valid dataset to show the recognition rate per user. The recognition rates were determined through a probability measurement performed by counting the words in a coherent text (e.g., a note). Confidence intervals were utilized to measure the uncertainty around estimates (Dix, 2020) and facilitated the assessment of technical affordances that either support or interfere with documentation work.

$$\text{Lower Limit:}\quad \frac{1}{n+z^2}\cdot\left(r+\frac{z^2}{2}-z\cdot\sqrt{\frac{r\cdot(n-r)}{n}+\frac{z^2}{4}}\right)$$

$$\text{Upper Limit:}\quad \frac{1}{n+z^2}\cdot\left(r+\frac{z^2}{2}+z\cdot\sqrt{\frac{r\cdot(n-r)}{n}+\frac{z^2}{4}}\right)$$

Equation 1: Confidence interval for proportion *p where n* = the number of words, *r* = the number of recognition errors, and *z* corresponds to the confidence level.

Samples were observed during ethnographic fieldwork that was conducted in 2022 for three months. Three HCPs using ASR were shadowed in outpatient clinics and wards at a Danish regional hospital. The study included patient consultations, while in situ interviews occurred (Gawlik, 2018; Czarniawska, 2021). Empirical data were organized into detailed descriptions (Baarts, 2015) and further analyzed iteratively to identify recurring themes, allowing for a comprehensive understanding of data work and emerging technical issues. Quotes are translated from Danish.

ANALYSIS

To improve recognition, HCPs train the ASR system by labeling mistranslated words during EHR documentation which serves as training data. HCPs often expressed uncertainty about the volume of labeling necessary to enhance recognition rates in the ASR system. They emphasized the need for ongoing training of the language model but were confused about why "they were responsible for understanding the system's functionality" (Interview, HCP). Besides them not feeling properly educated in using ASR, the confusion was underlined by the inability to access metrics that illustrated their recognition rates. This access denial left them unaware of whether their performance was improving or worsening ML whenever they labelled the ASR-produced output. Consequently, the time they spent on documentation—time that was allocated to patient care—was essentially repurposed by external suppliers, who relied on HCP expertise to update the language model which ASR draws on. Time that was already under pressure.

In contrast, medical secretaries and IT professionals had access to visualizations and metrics regarding HCP usage of ASR; however, these calculations were often overlooked. The metrics primarily illustrated unclear parameters of individual HCPs' ASR usage. The default recognition rate started at 100% and decreased each time an HCP corrected mistranslated words. Medical secretaries and IT professionals could only retrieve basic data on the hours HCPs had utilized speech recognition and the number of notes produced, accompanied by a corresponding percentage: "It may as well be 97%, but it is an expression of a physician who does not correct [label] much" (Interview, medical secretary). While these metrics served as performance indicators for evaluating HCPs' correction activities and contributed to enhancing the language model, secretaries found the monitoring tools inadequate.

Ignoring Glitches

During the observations of HCPs, several incidents occurred in which HCPs attempted to utilize ASR while encountering technical issues that caused the system to freeze upon activation. The nature of these glitches was unclear to the HCPs, leading them to attempt improvised solutions, such as unplugging all hardware components. Additionally, when software error pop-up messages stated that ASR was not functioning, HCPs often cancelled or minimized these messages or restarted the computer. HCPs lacked a clear understanding of the root causes of the errors and were unable to effectively communicate these issues to IT professionals when reporting them. They observed that delaying EHR documentation or using improvised methods sometimes enabled them to continue using ASR, even after encountering initial error messages.

Medical secretaries play a crucial role in the validation and management of data in the ASR system. Their responsibilities include validating the training data labeled by HCPs to ensure its accuracy, and then relaying this validated data to suppliers for updating clinical language models. This process is essential for maintaining the accuracy and functionality of the ASR system in healthcare settings. However, the secretaries faced challenges in holding HCPs accountable for proper ASR usage due to the lack of detailed quantitative assessments of HCP performance. Additionally, limitations in accessing certain areas of the ASR interface further hindered their ability to effectively validate the data entries made by HCPs, disrupting the overall workflow of maintaining data quality in the ASR system. Consequently, they prioritized other urgent tasks, neglecting glitches that prevented data from being forwarded to suppliers, ultimately hindering the suppliers' ability to provide a conversant clinical language model. The lack of these updates could have rendered the system outdated and less effective, particularly as labeled data was deleted after a few weeks and not reused for updates.

Interruptions

HCPs lacked access to trace their labelling which led to uncertainty around recognition rates. Table 1 shows the individual recognition rates of three HCPs who had trained the ASR system for over five years, highlighting the precision of the system in recognizing each HCP's speech. The table details the total labeling as corrections, total of words inputted in EHR, alongside the interval range expressed as a percentage. Notably, only one HCP achieved the anticipated recognition rate of at least 95%, underscoring the challenges in meeting this commercial benchmark.

Table 1. Confidence interval.

HCP	Total corrections	Total words	Interval start	Interval end	Interval	Interval percent
John	286	5326	0.04796011	0.0600809	[0.04796; 0.06008]	95.2%; 94%
Kasper	497	8625	0.05289938	0.06274087	[0.05290; 0.06274]	94.71%; 93.73%
Peter	425	5663	0.06847083	0.082202	[0.06847; 0.08220]	93.15%; 91.78%

The interruptions caused by ASR usage impact HCPs' workflow, while required to manage a high volume of patients during limited time. It is challenging for them to identify the amount of time using ASR and time spent to correct mistranslated words and labeling them. Consequently, determining when documentation work begins and ends becomes a complex and fluid process. In this context, Table 2 visualizes the number of corrections made by HCPs when documenting EHRs. During their shifts, John and Kasper documented 73 EHRs, and Peter 74. The table indicates an estimated average of corrections during shifts over a month ($n = 10$), revealing that documentation work were interrupted by an average of 602 corrections per month. This translates to over 40 interruptions per HCP during observed shifts.

Table 2. Corrections of mistranslated words.

HCP	Total corrections	Total words	Observed Shifts	Corrections in a month
John	286	5326	6	477
Kasper	497	8625	8	621
Peter	425	5663	6	708

Additionally, the official recognition rates calculated for Peter by the ASR system were inaccessible for over a year due to a glitch. He was not aware due to the missing accessibility. Secretaries and IT professionals do not overlook or do samples which could highlight such critical gap in system performance and support. Despite the system recognizing between 93.15% and 91.78% of Peter's spoken words, his profile was compromised, and unnoticed glitches persisted, revealing a limited awareness of individual ASR training development.

Further, the developer-reported metrics for ASR recognition rates for the department observed was cited as over 95%. This rate can be misleading because they included data from users who do not actively use the system or make minimal corrections although more were needed. These non-users or light users are effectively counted as having perfect (100%) recognition, as there is no opportunity for the system to misrecognize their input. This inflates the overall recognition rate, creating a discrepancy between the advertised performance and real-world results from active users.

DISCUSSION

To improve recognition rates, HCPs train the ML system by annotating mistranslated words during EHR documentation. This continuous annotating process effectively turns HCPs into human annotators extending their responsibilities beyond traditional clinical duties. Consequently, HCPs must balance documentation work with technical tasks, complicating their professional roles. Labeling is accompanied by significant challenges, primarily due to the lack of explainability and insufficient training causing HCPs to feel overwhelmed by the expectation to "understand" the system's data processing. To this extent, Bossen and colleagues (2019) emphasize the importance of HCPs having greater insights into data processing to not risking users to dodge these systems. Yet, despite hard-to-explain challenges, HCPs continued to use ASR and had done so for over five years. Further, they often struggle to identify and explain technical issues when communicating with IT professionals, making it difficult for them to effectively report problems. This lack of clarity in communication can lead to users perceive the systems as unhelpful or unreliable when malfunctions occur, ultimately reducing opportunities for collaborating training of ML algorithms. As Lin and Jackson (2023) emphasize, effective collaboration requires structured channels for users to share and document the specific technical issues they encounter. However, the current form of collaboration—a simple phone call to the local IT department—presents several challenges that hinder effective communication and problem-solving. These challenges include a lack of visual context, limited real-time interaction, and communication barriers between HCPs and IT professionals. As a result, technical issues are often prolonged, frustration grows, and trust in the system diminishes, undermining both collaboration and system improvements. If users cannot identify or understand technical issues, it greatly reduces their ability to effectively interact with the system. A key issue for HCPs using ASR systems is their limited access to recognition accuracy data. Since they are not assigned official data work *roles*, such as "transcription reviewers," they become overlooked in this process and lack the necessary feedback on how well the system performs, making it difficult for HCPs to engage meaningfully with the ASR system. This lack of clarity around the impact of their annotation efforts relegates HCPs to the role of passive users, further complicating their responsibilities and increasing their accountability aligning with how Bossen and Bertelsen (2023) calling for more formal recognition of clinical data work.

In the current empirical study, a technical issue prevented medical secretaries from accessing a critical tool for validating annotated data. This tool was essential for forwarding the data annotated by HCPs to system suppliers, and its absence disrupted the flow of data. This disruption ultimately affected language model updates and compromised the accuracy of the ASR system. Identifying and addressing technical glitches is vital to avoid what this paper terms "faux data work"—where seemingly productive efforts of data collection, management, analysis, and interpretation are rendered ineffective due to unresolved technical problems. For instance, documentation workflows were interrupted by hundreds of corrections each month, reflecting a burden of data work. Faux data work occurs when HCPs, lacking the necessary insights or metrics, cannot understand how their efforts, such as annotating or correcting data, impact a system. Consequently, a glitch in an ASR system is not just a minor issue but reflects deeper systemic problems. HCPs continue using these systems despite flaws, often for years, without access to key metrics like recognition accuracy. This prolonged use, without addressing technical issues, suggests a normalization of system faults. HCPs, experiencing frequent corrections and workflow disruptions, still rely on the system, which reduces the effectiveness of their work and risks entire departments relying on flawed a data lifecycle.

Mitigate Faux Data Work

To effectively address the issue of faux data work, two key strategies are proposed.

First, it is essential to provide HCPs with transparent and actionable system metrics. Currently, HCPs often lack access to critical feedback mechanisms, such as recognition accuracy and the impact of their annotation efforts, which are necessary for them to assess the effectiveness of their contributions to the system. The provision of clear, real-time metrics, such as recognition rates and annotation progress, would enable HCPs to monitor their own and collaborative performance and make informed decisions regarding their engagement with the system. Transparency would help HCPs avoid inefficient tasks, thereby reducing the occurrence of faux data work. By empowering HCPs with the ability to gauge the influence of their actions, this approach enhances their ability to contribute meaningfully to the system's functionality.

Second, enhancing collaboration between HCPs and IT professionals through structured sites of collaboration is suggested. The current reliance on phone calls for troubleshooting technical issues can result in miscommunication or delays in issue resolution. Building on Lin and Jackson's (2023) advice for platforms that facilitate real-time reporting of technical issues, utilizing visual aids such as screenshots or screen recordings could further improve the speed and accuracy of issue resolution. The introduction of guided troubleshooting templates and visual communication channels would assist technical problems to be addressed promptly (Yao *et al.*, 2024), preventing the persistence of glitches and reducing the risk of faux data work within healthcare workflows.

CONCLUSION

Despite the critical role HCPs play in interacting with ASR, there remains limited knowledge about how they detect and address technical issues within these systems. If HCPs do not fully comprehend the data processing and system outputs, they risk inadvertently training ML incorrectly, leading to "faux data work," where seemingly productive efforts fail to yield meaningful results. To prevent this waste of time and efforts, two critical improvements are necessary: providing clear, interpretable metrics that allow HCPs to track their labeling performance, and enhancing explainability to equip HCPs with the knowledge needed to navigate and troubleshoot ASR.

REFERENCES

Adedeji, A., Joshi, S. and Doohan, B. (2024) 'The Sound of Healthcare: Improving Medical Transcription ASR Accuracy with Large Language Models'. arXiv.

Aldarmaki, H. et al. (2022) 'Unsupervised Automatic Speech Recognition: A review', Speech Communication, 139, pp. 76–91.

Baarts, C. (2015) 'Introduktion til etnografisk metode: Metodeserie for social-og sundhedsvidenskaberne'.

Baevski, A. et al. (2022) 'data2vec: A General Framework for Self-supervised Learning in Speech, Vision and Language'. arXiv.

Bertelsen, P. S. et al. (2024) 'Data work and practices in healthcare: A scoping review', International Journal of Medical Informatics, 184, p. 105348.

Bi, N. and Huang, J. Y.-C. (2023) 'I create, therefore I agree: Exploring the effect of AI anthropomorphism on human decision-making', in CSCW '23: Computer Supported Cooperative Work and Social Computing, Minneapolis MN USA: ACM, pp. 241–244.

Bossen, C. et al. (2019) 'Data work in healthcare: An Introduction', Health Informatics Journal, 25(3), pp. 465–474.

Bossen, C. and Bertelsen, P. S. (2023) 'Digital health care and data work: Who are the data professionals?', Health Information Management Journal.

Chen, S. et al. (2022) 'WavLM: Large-Scale Self-Supervised Pre-Training for Full Stack Speech Processing', IEEE Journal of Selected Topics in Signal Processing, 16(6), pp. 1505–1518.

Czarniawska, B. (2021) 'How to shadow organizing', in Organizational Ethnography. United Kingdom: Edward Elgar Publishing Limited. Available at: https://doi.org/10.4337/9781786438102.00008.

Dix, A. (2020) 'Statistics for HCI: Making Sense of Quantitative Data', Synthesis Lectures on Human-Centered Informatics, 13(2), pp. 1–181.

Gawlik, K. (2018) 'Focus group interviews', in Qualitative methodologies in organization studies. Springer, pp. 97–126.

Hacking, C. et al. (2023) 'The development of an automatic speech recognition model using interview data from long-term care for older adults', Journal of the American Medical Informatics Association, 30(3), pp. 411–417.

Haque, M. R. and Rubya, S. (2022) "'For an App Supposed to Make Its Users Feel Better, It Sure is a Joke"—An Analysis of User Reviews of Mobile Mental Health Applications', Proceedings of the ACM on Human-Computer Interaction, 6(CSCW2), pp. 1–29.

Herrmann, T. and Pfeiffer, S. (2023) 'Keeping the organization in the loop: A socio-technical extension of human-centered artificial intelligence', AI & Society, 38(4), pp. 1523–1542.

Kaswan, K. S. et al. (2021) 'AI-Based Natural Language Processing for the Generation of Meaningful Information Electronic Health Record (EHR) Data', in Advanced AI Techniques and Applications in Bioinformatics. CRC Press.

Kumar, Y. (2024) 'A Comprehensive Analysis of Speech Recognition Systems in Healthcare: Current Research Challenges and Future Prospects', SN Computer Science, 5(1), p. 137.

Lee, J. D. and Kolodge, K. (2020) 'Exploring Trust in Self-Driving Vehicles Through Text Analysis', Human Factors, 62(2), pp. 260–277.

Lewis, J. R. (2016) Practical Speech User Interface Design. CRC Press.

Lin, C. K. and Jackson, S. J. (2023) 'From Bias to Repair: Error as a Site of Collaboration and Negotiation in Applied Data Science Work', Proc. ACM Hum.-Comput. Interact., 7(CSCW1), pp. 131:1–131:32.

Lott, A. and Reiter, J. P. (2020) 'Wilson confidence intervals for binomial proportions with multiple imputation for missing data', The American Statistician, 74(2), pp. 109–115.

Meripo, N. and Konam, S. (2022) 'ASR Error Detection via Audio-Transcript entailment', in, pp. 3358–3362.

Ngueajio, M. K. and Washington, G. (2022) 'Hey ASR System! Why Aren't You More Inclusive?', in J. Y. C. Chen et al. (eds) HCI International 2022 – Late Breaking Papers: Interacting with eXtended Reality and Artificial Intelligence. Cham: Springer Nature Switzerland, pp. 421–440.

Palivela, H., Narvekar, M. and Tiwari, N. (2023) 'Responsible AI in Automatic Speech Recognition', in Era of Artificial Intelligence. Chapman and Hall/CRC.

Pascual, R., Ing, J. A. and Azcarraga, J. (2024) 'TDNN-HMM ASR Systems on Under-Resourced Local Languages Towards Application in a Healthcare Chatbot', in V. Thiruchelvam et al. (eds) Proceedings of the 4th International Conference on Advances in Computational Science and Engineering. Singapore: Springer Nature, pp. 469–478.

Raji, I. D. et al. (2022) 'The Fallacy of AI Functionality', in Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency. New York, NY, USA: Association for Computing Machinery (FAccT '22), pp. 959–972. Available at: https://doi.org/10.1145/3531146.3533158.

Terriza, M. et al. (2022) 'Use of Laughter for the Detection of Parkinson's Disease: Feasibility Study for Clinical Decision Support Systems, Based on Speech Recognition and Automatic Classification Techniques', International Journal of Environmental Research and Public Health, 19(17), p. 10884.

Turri, V. and Dzombak, R. (2023) 'Why We Need to Know More: Exploring the State of AI Incident Documentation Practices', in Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society. AIES '23: AAAI/ACM Conference on AI, Ethics, and Society, Montr\'{e}al QC Canada: ACM, pp. 576–583.

Vase, S. and Berget, G. (2023) 'An Exploration of Automatic Speech Recognition Within a Nordic Context', in M. Antona and C. Stephanidis (eds) Universal Access in Human-Computer Interaction. Cham: Springer Nature Switzerland (Lecture Notes in Computer Science), pp. 288–307.

Yao, Z. et al. (2024) 'Do Physicians Know How to Prompt? The Need for Automatic Prompt Optimization Help in Clinical Note Generation'. arXiv.