doi.org/10.58931/cdt.2025.63147

ABOUT THE AUTHOR

Juthika Thakur, MD

Dr. Juthika Thakur completed her Bachelor of Medical Sciences degree from Western University and a business degree from the Richard Ivey School of Business in 2011 with honours. She then went on to graduate from Michael G. DeGroote School of Medicine at McMaster University and completed her dermatology residency at the University of Toronto serving as a co-chief resident in her final year. Since then, Dr. Thakur has written and presented her research at several national and international conferences such as, the Canadian Dermatology Association, the European Academy of Dermatology and Venereology, The World Congress of Dermatology, and is published in the Ivey Business Review. She has an interest in the intersection between e-health, machine learning, and dermatology.

A Dermatologist's Beginner's Guide: Using Ambient Artificial Intelligence in Your Practice

Juthika Thakur, MD

Introduction

Automatic Speech Recognition (ASR) is the fundamental technology that enables the conversion of spoken language into written text. The strengths of ASR include interpreting voices, identifying different speakers in a conversation, and following dialogue. With the support of human trainers, ASR systems can further improve and optimize their performance based on feedback. However, their effectiveness can be limited by factors such as noisy environments, poor microphone placement, and variations in dialects and accents. See **Exhibit 1** for definitions of key AI terms.

Large Language Models (LLMs) aim to bridge these gaps by using Natural Language Processing (NLP) to fill in missing elements of a conversation. As a type of generative AI, an LLM predicts the next word in a sequence based on patterns found in its training dataset (**Figure 1**). The goal of an LLM is to mimic human language by identifying which words are likely to follow one another, given the context provided in the prompt and its prior training. LLMs can generate original content and optimize their outputs based on human input. However, because they function by predicting word sequences, they do not inherently understand whether their outputs are true or false. As a result, LLMs can produce inaccurate or fabricated responses, commonly referred to as "hallucinations" in their responses to prompts.

Real-time ambient AI scribes, which leverage machine learning to process conversations, show promising potential to reduce the documentation burden, enhance the quality of doctor–patient interactions, and support clinicians in their daily

Key AI Terms and Definitions
Artificial Intelligence: A field of computer science focused on creating systems that can perform tasks normally requiring human intelligence, such as reasoning, problem-solving, learning, and language understanding.
Machine Learning: A subset of AI where computers learn patterns from data to make decisions or predictions without being explicitly programmed for each task.
Augmented intelligence: A collaborative approach where humans and AI systems work together to enhance human decision-making and performance, rather than replacing human intelligence.
Agentic AI: AI systems designed to operate autonomously and pursue goals or perform actions proactively, often with some level of decision-making or initiative.
Large Language Models: AI models trained on vast amounts of text data to understand and generate human-like language. Examples include ChatGPT, Claude, and Gemini.
Natural Language Processing: A branch of AI focused on enabling machines to understand, interpret, and generate human language, both written and spoken.
Hallucinations: When an AI model generates information that sounds plausible but is actually false or not based on real data.
Supervised learning: A type of machine learning where the model is trained on labelled data (input-output pairs), allowing it to learn to make predictions or classifications.
Unsupervised learning: A machine learning technique where the model is trained on data without labelled outcomes, often used to find patterns or groupings (like clustering).
Bias: Systematic errors in AI outputs caused by imbalanced or flawed training data, design choices, or societal inequalities, leading to unfair or inaccurate results.
Explainability and Interpretability: The degree to which humans can understand how an AI model works, including how it makes decisions or predictions. Crucial for trust and transparency in AI systems.

Exhibit 1. Key AI Terms and Definitions; *courtesy of Juthika Thakur, MD.*

work. When combined, LLMs and ASRs can offset each other's limitations, resulting in an ambient AI scribe that is more effective and practical for use in clinical settings (**Figure 2**).

Accuracy of the Generated Electronic Medical Record by Physician Versus Ambient AI

Clinical documentation may not always fully capture the substance of the patient encounter, with omissions that do not necessarily reflect the nature or intent of the interaction. In a study that included 36 physicians at a single centre, written encounters were compared with concealed audio recordings in unannounced patient encounters. The findings showed that 90% of patient records had at least one inaccuracy, with omissions and errors of additions that did not reflect the nature of the interaction captured in the concealed audio recordings.[1] In another study, among 136,815 patients who reviewed their outpatient clinic visit notes, 20% detected an error, and 40% of those considered the error to be serious.[2] The most common errors were in diagnoses, medical histories, medications, physical examinations, and misattributed notes to the wrong patient.[2] Many users of AI complain about the diagnostic accuracy of machine learning algorithms and LLMs. Unfortunately, the patient chart often falls short as an accurate reflection of the patient-physician encounter. In a regional pilot program that scaled AI based medical scribe solutions to 10,000 physicians, both patient and physician experiences improved.[3] A random review of 35 assessed notes across multiple clinical specialties revealed that over 90% met metrics such as freedom from hallucinations, conciseness, and accuracy (**Exhibit 2**). Although this study showed nominal improvements in charting time,[3] other studies have shown a reduction in time spent charting by 20.4%.[4] Domain-specific

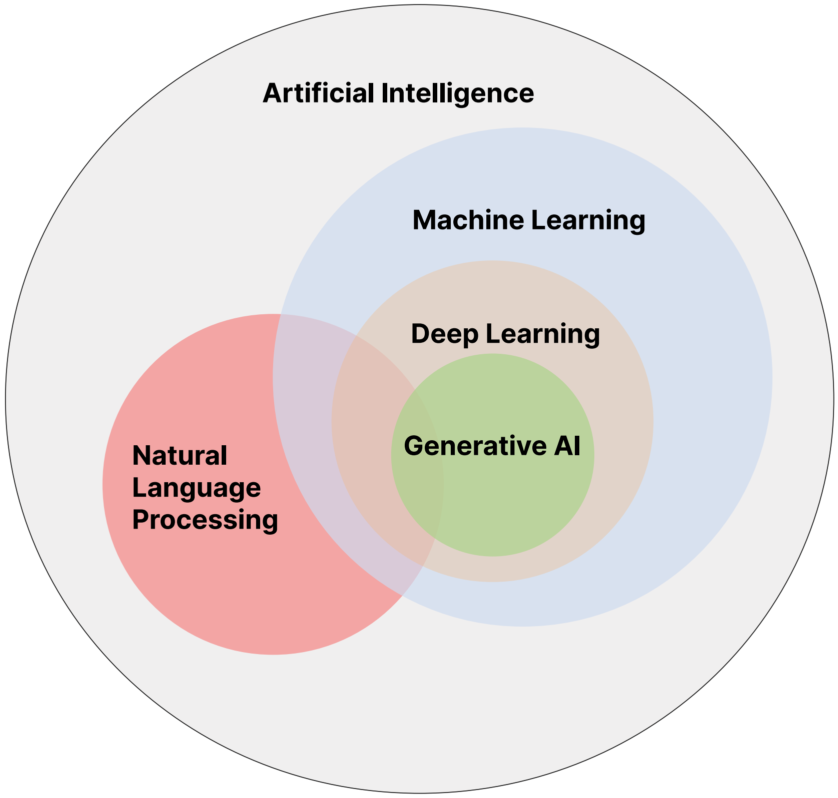

Figure 1. A Venn diagram depicting various subdomains of AI; *adapted from Rishabh Misra, 2024.*[7]

Automatic Speech Recognition

Basic Capabilities

- ✓ Capture, recognize, and interpret voices
- ✓ Identify different people in conversations
- ✓ Follow conversational patterns

Basic Limitations

- → Susceptible to environmental conditions – placement of microphones, background noise
- → May struggle with terms, accents, or dialects not part of its training data
- → May struggle with fragmented and non-linear conversations

Large Language Models

Basic Capabilities

- ✓ Identify words likely to go together given training data and context from the prompt
- ✓ Produce realistic and persuasive outputs
- ✓ Generate original content
- ✓ Learn from corrections of human users and trainers
- ✓ Improve and optimize performance over time

Basic Limitations

- → Quality of input determines the quality of output
- → Not designed to know if outputs are true or false
- → LLMs ad lib and hallucinate
- → Will struggle with concepts not part of its training data
- → Will parrot biases present in training data

Figure 2. Automatic Speech Recognition and Large Language Models Strengths and Weaknesses; *adapted from Information Services Centre of Effective Practice. (Anne Dabrowski, n.d.)*[8]

Attribute	Description of Ideal Note
Accurate	The note is true. It is free of incorrect information.
Thorough	The note is complete and free from omission and documents all of the issues of importance to the patient.
Useful	The note is extremely relevant, providing valuable information and/or analysis.
Organized	The note is well-formed and structured in a way that helps the reader understand the patient's clinical course.
Comprehensible	The note is clear, without ambiguity or sections that are difficult to understand.
Succinct	The note is brief, to the point and without redundancy.
Synthesized	The note reflects the AI scribe's understanding of the patient's status and ability to develop a plan of care.
Internally Consistent	No part of the note ignores or contradicts any other part.
Free from Hallucination	The note is free of hallucination and only contains information verifiable by the transcript.
Free from Bias	The note is free of bias and contains only information verifiable by the transcript and not derived from characteristics of the patient or visit.

Exhibit 2. Elements of an Ideal Clinical Encounter Summary; *adapted from Tierney et al., 2024.*[3]

A modified version of Physician Documentation Quality Instrument adapted for evaluating AI scribe outputs. Maximum value is 50, and each domain is based on a 5-point scale with 1 being not at all, and 5 being extremely likely.

training, with "humans in the loop" can mitigate against transcription inaccuracy. For example, by providing feedback and edits into the ambient AI module, outputs can be tailored to reflect the style of the clinician's encounters. Models that are not trained on healthcare-specific data are more likely to inaccurately capture healthcare lexicon, misinterpret drug names, inaccurately capture the complexities of medical conversations, and generally struggle with poor technology and noisy environments.

The Black Box, Explainability and Interpretability

Assessing diagnostic accuracy in LLMs remains challenging due to the opaque nature of their training datasets, which are often proprietary and not accessible to researchers. This lack of opacity limits the ability to identify, predict, or mitigate potential blind spots and biases in model outputs. Explainability refers to how clearly the processes behind an AI's output can be understood or justified. Effort has been made to put measures in place to provide some degree of explainability. For example, some systems include features such as the output note having "citations" back to specific segments of the audio transcript that the AI scribe summarized in a sentence format. This can mitigate against hallucinations but still requires intense physician effort to review and edit notes. Emerging research on physicians' use of AI indicates a genuine risk of automation bias, potentially leading to less thorough reviews and overlooking errors.[5] Ultimately, ambient AI vendors must strive to create technology that is explainable to all stakeholders, including patients and physicians on the use of AI scribes in medical encounters.

Privacy and Informed Consent with Using Ambient AI Technology

Clinicians have both ethical and legal obligations to obtain informed consent from patients. Patients may want to know information about how long recordings are stored, who can access the recording or AI summary, whether and how the physician will review them, whether the material collected is being used to train generative AI, and the potential risk of reidentification. A key challenge is that once patients have given consent, they cannot retract it for future use of the data in model training. As a result, informed consent should also include a discussion on whether and how data will be deidentified and used to improve the algorithm. To mitigate these risks, physicians can select an ambient AI scribe that hosts the data on local servers and avoids using patient data in perpetuity.

Challenges in Record Keeping

Although the final chart entry must be retained per record retention rules, laws and regulations typically do not specify whether such audio recordings should also be included in the patient's chart. When recordings are intended to be destroyed rather than stored, it is important to have a policy that dictates the timing and process for destruction, making sure the patient chart has been accurately updated beforehand.[6] Some regulatory colleges require an additional consent form specifically for recording clinical encounters, and in some provinces, privacy laws necessitate written consent for any recordings to take place.[6] Physicians should carefully review applicable privacy laws and provincial regulations before implementing AI scribe technologies into their workflows.

Summary

Methods for thoroughly assessing the quality and safety of AI technologies—including LLMs—are still not fully established. As both the algorithms and regulatory frameworks continue to evolve, continuous benchmarking, evaluation, and monitoring will be required. Additionally, adoption and usage are likely to shift as new user groups and application areas emerge.

Correspondence

Juthika Thakur, MD
Email: juthika.thakur@medportal.ca

Financial Disclosures

J.T.: None declared.

References

1. Weiner SJ, Wang S, Kelly B, Sharma G, Schwartz A. How accurate is the medical record? A comparison of the physician's note with a concealed audio recording in unannounced standardized patient encounters. J Am Med Inform Assoc. 2020;27(5):770–775. doi:10.1093/jamia/ocaa027
2. Bell SK, Delbanco T, Elmore JG, Fitzgerald PS, Fossa A, Harcourt K, et al. Frequency and types of patient-reported errors in electronic health record ambulatory care notes. JAMA Netw Open. 2020;3(6):e205867. doi:10.1001/jamanetworkopen.2020.5867
3. Tierney AA, Gayre G, Hoberman B, Mattern B, Ballesca M, Kipnis P, et al. Ambient artificial intelligence scribes to alleviate the burden of clinical documentation. NEJM Catal Innov Care Deliv. 2024;5(3). doi:10.1056/CAT.23.0404
4. Duggan MJ, Gervase J, Schoenbaum A, Hanson W, Howell JT, 3rd, Sheinberg M, et al. Clinician experiences with ambient scribe technology to assist with documentation burden and efficiency. JAMA Netw Open. 2025;8(2):e2460637. doi:10.1001/jamanetworkopen.2024.60637
5. Khera R, Simon MA, Ross JS. Automation bias and assistive AI: risk of harm from AI-driven clinical decision support. Jama. 2023;330(23):2255–2257. doi:10.1001/jama.2023.22557
6. CMPA. AI Scribes: answers to frequently asked questions Ottawa, Ontario, Canada: CMPA; 2023: [Cited June 6 2025] Available from: https://www.cmpa-acpm.ca/en/advice-publications/browse-articles/2023/ai-scribes-answers-to-frequently-asked-questions.
7. Misra R. What is generative AI? : Medium; 2024: [Cited June 6 2025] Available from: https://medium.com/analytics-vidhya/what-is-generative-ai-9c34cfa0fd6b.
8. Dabrowski A. OMD Educates—AI 101: Intro to AI and AI Scribe [Video Recording]: OntarioMD; 2024: [Cited March 22 2025] Available from: https://vimeo.com/925848578.