At Startup That Says Its AI Writes Medical Records, Humans Do a Lot of the Work

DeepScribe has a team of people to correct terminology, remove erroneous prescriptions and add billing codes

By *Joseph Walker* [Follow]

July 23, 2023 7:00 am ET

Hundreds of doctors across the U.S. have entrusted recordings of their private talks with patients to a startup promising to turn the conversations into usable medical records through artificial intelligence.

The technology makes multiple errors while producing the reports, such as failing to use correct medical terminology and adding medicines a patient isn't taking, according to current and former workers.

To fix those errors, health-tech startup DeepScribe relies on 200 human contractors to listen to the medical conversations and revise the records, the company's founders said. The workers also use Google searches to find billing codes.

With all its administrative tasks, data entry and paperwork, healthcare is a ripe target for AI. DeepScribe has made sweeping claims about the power of its AI, but the crucial work performed by its contractors shows the technology still can't pull off some basic chores in medicine without heavy human assistance.

Errors such as listing medications a patient hasn't taken are common to many AI algorithms, which often suffer from "hallucinations" where they present false information as true, DeepScribe co-founders Matthew Ko and Akilesh Bapu said.

Akilesh Bapu, left, and Matthew Ko co-founded DeepScribe in 2017. PHOTO: BUSINESS WIRE/ASSOCIATED PRESS

DeepScribe's software is able to write 80% of each record, the founders said, but the company employs 200 people to "catch errors that can be a product of nascent technology such as AI and give healthcare providers the confidence of knowing that their notes are reviewed by a training member of our team for accuracy."

Having humans review the AI's work and identify its mistakes helps train the algorithm and improve its accuracy, Ko and Bapu said. DeepScribe discloses the quality-assurance work on its website and in sales presentations, they said.

Doctors are responsible for ensuring the accuracy of the reports they add to patients' medical records. Current and former DeepScribe workers said they weren't aware of doctors failing to catch any errors that get through the human review.

The following account of DeepScribe, its technology and use of human workers is based on interviews with current and former workers, as well as a review of communications among the DeepScribe workers and other company materials.

The company works in a tedious but crucial corner of healthcare: record-keeping.

AI-based solution

Modern Technology for a Modern Practice

Artificial intelligence is revolutionizing virtually every industry. Now, it's healthcare's turn.

DeepScribe's ambient scribe technology extracts medical information from the natural conversation between you and your patient, summarizes the interaction into a complete note, then integrates that note into the appropriate fields of your EHR.

More time building relationships with patients. Less time documenting.

DeepScribe describes on its website how it employs AI to fill out electronic health records.

Doctors enter notes on their conversations into patient medical records. The doctors and other caregivers can later review the summaries to quickly get up to speed on a patient's medical history and help inform treatment. Health insurers use the records to determine how much to pay doctors for their services.

But writing up the chats into coherent notes is time-consuming work. To save time, many doctors simply copy and paste the notes they entered for the patient's previous visit and tweak only as needed. Some doctors will dictate patient reports using speech-recognition software or have a staffer accompany them on patient visits and take notes.

DeepScribe has made bold promises about the power of its technology to automate the record-keeping, saying it will "rewrite how medical documentation is done."

The company says it doesn't simply transcribe a doctor's notes of a patient visit. Its AI takes verbatim transcripts of each conversation, which can total thousands of words, and turns them into standardized reports that can be merged into the patient's electronic medical record.

DeepScribe provided images from its sales presentation and the FAQ section of its website that include disclosures that a quality-assurance team reviews the AI-generated reports for accuracy.
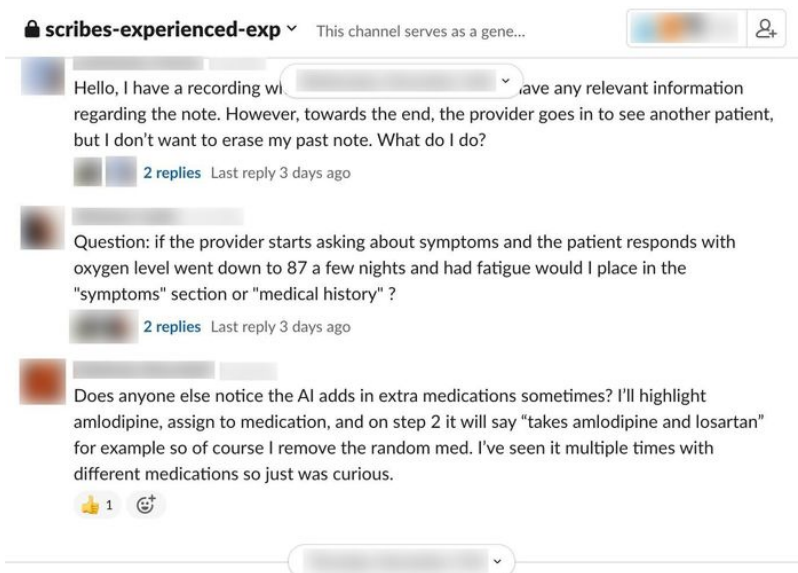
Yet most of the website emphasizes the benefits of its "AI-powered" tech. The software is "more accurate" and "less manual" than older services, the company says on the site. "Artificial intelligence is revolutionizing virtually every industry. Now, it's healthcare's turn," the site says.

Healthcare spending accounts for nearly one-fifth of U.S. GDP, making it one of the hottest and most potentially lucrative markets for AI. Due to the commercial potential, venture capitalists invested $17.3 billion in healthcare-focused AI companies from 2018 to 2022, according to PitchBook, a data analytics provider.

For many health-tech startups, however, there is a wide gap between the promises they make about their technology and the reality of what they deliver, said Dr. Isaac Kohane, professor and chairman of biomedical informatics at Harvard Medical School.

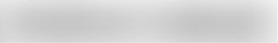"The froth is such that there is too much money chasing too few good companies," he said.

San Francisco-based DeepScribe, founded in 2017, has raised $37.3 million in venture-capital financing, and was valued at $180 million after its last financing in December 2021, according to PitchBook.

The company has about 1,000 doctors and other healthcare providers using its service, Ko and Bapu said. They said the company was on track to have revenue of substantially more than $6.5 million this year.

The company charges one-sixth to one-eighth of the $2,000 to $3,000 that other AI doctors' notes competitors bill a medical practice each month, the founders said.

An employee polled a DeepScribe Slack channel for advice on terminology.

With DeepScribe's product, doctors record each patient visit with a smartphone application made by the company, and then upload the recording to DeepScribe's database.

DeepScribe uses software from other companies to produce a transcript of the audio, Ko and Bapu said. Then DeepScribe's AI gets to work summarizing the transcript for placement in a patient's medical record.

The AI analyzes the transcript and highlights medically relevant information, such as a patient's medical history and current symptoms, as well as the doctor's follow-up plan. The AI writes up the highlighted information in complete sentences.

DeepScribe's human workers, known internally as scribes, review the audio recordings and compare them with the transcripts and AI-generated sentences.

The problems start with the transcripts, the current and former workers said. Often they contain disjointed, nonsensical sentences.

DeepScribe says its software also recommends key pieces of data to be included in patients' medical records, including codes for billing insurance.

Both the transcripts and the AI-generated summaries also sometimes incorrectly spelled the names of medications, the workers said. For instance, one transcript referred to a drug that a patient said he was taking for migraines as "Shelby," when the patient had actually used the brand-name Ubrelvy, one of the people said.

DeepScribe software sometimes suggested changes to a medicine that was correctly spelled in the transcript. For example, it recommended changing the name of the antidepressant Remeron to Remora, a type of fish, or proposing Toreador, a Spanish word for bullfighter, in place of pain reliever Toradol, one person said.

"Does anyone else notice the AI adds in extra medications sometimes?" one worker reviewing the AI's work wrote in a Slack message to colleagues in November 2022 that was reviewed by The Wall Street Journal. "I've seen it multiple times with different medications so just was curious."

The current and former workers said they frequently had to make corrections and fill in other gaps left by the AI. They said they often composed most of each medical report.

The Journal's review of messages on the Slack channel used by scribes found them seeking answers to a range of questions. Among the topics discussed were whether to describe an unnamed patient, who complained to a doctor about not sleeping well, as suffering from sinus congestion or as having insomnia.

DeepScribe's AI often failed to replace the colloquial language spoken by patients with medical terminology, Slack messages viewed by the Journal show. The company expected its human employees to perform that function.

"Is there a medical word for belly button lol," one scribe asked. "Navel!" a DeepScribe administrator responded. "Or Google is also telling me 'umbilicus.'"

The quality of the transcripts and reports has improved over the past several months after the company began using a new version of its AI software, one of the workers said.

Ko and Bapu said the algorithm also learns each doctor's conversation and writing styles, and improves its accuracy. Employees said that they were responsible for making the notes conform to doctors' preferences, which were communicated by doctors to DeepScribe's sales staff.

DeepScribe said its AI constructs the bulk of the reports and its human workers scan transcripts and reports for any information that the AI may have missed. The workers raise the accuracy of the AI's work by 15 percentage points to 95%, and the remaining 5% is done by doctors proofing the notes before accepting them, Ko and Bapu said. Doctors contacted by the Journal who use DeepScribe's software didn't respond to interview requests.

The scribes also were responsible for selecting the billing codes used to identify each patient's diagnoses, symptoms and procedures because DeepScribe's AI isn't able to retrieve the codes, the current and former workers said.
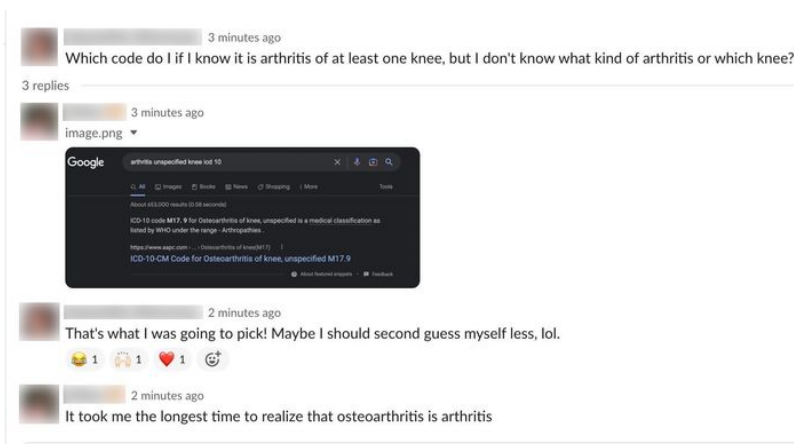
Employees on a DeepScribe Slack channel crowdsourced advice on which medical billing codes to include in notes.

To do their coding, many large medical practices and hospitals employ specialized staff who have completed training in medical coding through certificate or associate degree programs. DeepScribe's training of scribes was minimal, and included specifying the common codes to use and how to search the internet for others, one of the workers said. Managers instructed scribes to use Google to find the diagnosis, or ICD-10, codes, the people said.

Early this year, a scribe asked on Slack which ICD-10 code to use for knee arthritis without knowing what kind of arthritis or which knee. A manager responded with a screenshot of a Google results page for the search terms "arthritis unspecified knee icd 10."

DeepScribe said it provides scribes training in ICD-10 and that it only recommends the codes to doctors, who are responsible for confirming their accuracy.

Some doctors and patients appeared to be unaware that humans are listening to the recordings, the current and former workers said. On some recordings, doctors told patients that their visit was being recorded, but would be analyzed entirely by AI and that no people would listen to it, according to the workers.

During the conversations, patients have shared intimate details with their physicians, including cases of drug abuse and family problems, the DeepScribe workers said. The transcripts often include the patient's full name and an abbreviation of the doctor's practice name, the people said.

Under the federal Health Insurance Portability and Accountability Act, or HIPAA, DeepScribe and its employees are permitted to review patients' medical information but they are bound to safeguard the confidentiality of the data, legal experts said.

The current and former workers said they received minimal training in the federal HIPAA law and health privacy generally.

Legal experts said HIPAA doesn't require doctors to tell patients that other people may examine their data, including recordings of their visits.

Yet even if legal, doctors for ethical reasons should tell patients that their conversations are being recorded and shared with a third-party company and its employees, said Joy Pritts, a health-privacy consultant to tech companies.

"They're misrepresenting what they're doing to the patient, and in a way which makes the patient feel more comfortable and freer about what they say," said Pritts.

DeepScribe said all employees complete a 17-step training program, which includes HIPAA compliance.

Write to Joseph Walker at joseph.walker@wsj.com

Appeared in the July 24, 2023, print edition as 'Medical-Record Startup Uses AI, Human Power'.