How Private Is Medical Speech Data?

Not very, as 'anonymized' data and privacy policies show. Meaningful consent might be a start

Data is currency. It has value. Its use or misuse can result in privacy breaches. Individuals should control how their data is being used and when and how—or if—they want to share it.

Speech recognition vendors iScribe, M*Modal, Nuance Communications, and Speech Processing Solutions (SPS) have varying opinions as to who owns customer data. SPS does not use dictation data for any purpose other than to process the dictation document.

iScribe, M*Modal, and Nuance, however, assume ownership rights to customer dictation data: "In the event that Personal Data are to be used for a new purpose incompatible with the purposes for which the data were originally collected by M*Modal's customers...M*Modal's customers will be given notice of such so that customers can provide notice to data subject and, **where feasible and appropriate**, an opportunity to decline to have their data so used or transferred." [Emphasis added.]

Right. Because doctors don't have enough to do as it is. Now, apparently, they are tasked with keeping their patients current about all the ways M*Modal could be using dictation data. And what does incompatible mean? If a doctor thinks she is dictating for the sole purpose of having her dictation data translated into a text document, any other use is incompatible.

By using Nuance products "you consent to the collection and use of your Personal Information by Nuance.... You also represent that you have any and all authorizations necessary to use these Nuance Products including using them to process Personal Information. Nuance collects and uses the information you provide...**for our internal purposes**." [Emphasis added.]

This includes "data collection" of customers' dictation data—going so far as to hijack user data from single-license desktop installations, which "defeats the whole purpose of a [local install]," says attorney David Schwartz, "because absent assent, there is an expectation of privacy." And in such a case, the expectation is false.

"What's in the dictation?" asks attorney Lynne Geminder. "My voice, my thoughts, my ideas, my conversations. The only thing that's theirs is the server—and I never gave them permission to put it there in the first place."

The Anonymization Myth

M*Modal's policy states that "personal data does not include aggregate data that is not individually identifiable," which likely means that M*Modal and iScribe aggregate data.

The claim that data anonymization should allay customer concerns regarding privacy falls flat. Where companies retain data for what could be forever, the privacy policy landscape is disrupted because companies rely on anonymization to justify their actions, advancing the misleading appearance that they are protecting privacy. "Advances in reidentification expose these promises as too often illusory," writes Paul Ohm in his 2010 article "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization."

In 2006, Netflix publicly released a large data set of customer movie ratings and preferences after claiming that all personally identifiable information (PII) had been removed. A researcher was easily able to de-anonymize the data.

"Scientists have demonstrated that they can often... 'deanonymize' individuals hidden in anonymized data with astonishing ease," writes Ohm, calling into question the whole paradigm of protecting privacy by removing PII.

Meaningful Consent

Privacy policies are not the same as informed consent, with "privacy policies being more concerned about institutional liability than individual well-being," says ethicist Anna Lauren Hoffman.

Privacy policies derive from contract law, while informed consent is born from often protracted negotiations among parties. Individual users are typically not consulted and are often unaware of associated risks.

Hoffman reminds us that Big Data is about people, not numbers. When a company is extracting data from anything—whether it's a wearable device or a cloud-based speech recognition server—it is easy to lose sight of the people on the other end. Data is "intimately bound up with individual human lives," Hoffman says.

"If material information regarding consent is buried in a set of terms and conditions, that in itself is a problem," says Marty Abrams, executive director of the Information Accountability Foundation.

Users should have meaningful control over data that pertains to them, whether they're doctors, patients, or third parties referenced in the dictation.

Data should be collected based on appropriate consent, with users understanding what they're consenting to and the risks involved. Companies should ask (1) whether the use/processing of the data is fair; (2) whether the individual understands the impact of the data use; (3) whether the use creates value for the individual; and (4) whether the use/processing is detrimental to the individual. ☒

Robin Springer is an attorney and the president of Computer Talk, Inc. (www.comptalk.com), a consulting firm specializing in speech recognition and other hands-free technology services. She can be reached at (888) 999-9161 or contactus@comptalk.com.