

Brendan R.

brreed@andrew.cmu.edu

94-806 Privacy in the Digital Age

Dr. Joanne Peca

Homework 3: Readings Critique

Narayanan and Shmatikov's 2010 article "Myths and Fallacies of 'Personally Identifiable Information'" published in Communications of the ACM represents a challenge of traditional privacy legislation (or lack thereof) on the concept of personally identifiable information (PII). Arvind and Vitaly's critique express that current anonymization techniques do not sufficiently deidentify an individual, especially enough to remain anonymous upon the joining of two "anonymized" datasets. In an increasingly digital age where individual's data is in the hands of countless brokers and data stewards, the idea of a comprehensive elimination of PII is an elusive idea which no organization can completely claim without adopting further privacy techniques that go beyond current laws and regulations.

Narayanan and Shmatikov's 2010 critique¹ remain largely valid. However, their argument that PII is "context-dependent" potentially underestimates the degree to which some data elements are substantially more identifying than others. While Social Security numbers, email addresses, and biometric data do not identify in all contexts, they are datapoints with significantly more weight than attributes like age, zip code, or fax number.² Contemporary guidance, such as NIST Special Publication 800-122, acknowledges this distinction by ranking PII according to potential impact (low, moderate, high) in the event of disclosure, rather than treating all identifiers equally. This suggests that while context matters, not all data elements contribute equally to re-identification risk.³

Second, the article's conclusions and recommendations are somewhat limited. After dismantling the current PII landscape, Narayanan and Shmatikov offer relatively little guidance on what should replace it. They gesture toward approaches like differential privacy but do not engage deeply with the practical challenges of implementation⁴. Even

¹ Narayanan, A., & Shmatikov, V. (2010). Myths and fallacies of "personally identifiable information." *Communications of the ACM*, 53(6), 24–26. <https://doi.org/10.1145/1743546.1743558>

² The 18 HIPAA identifiers | Information Technology Services (ITS): Loyola University Chicago. (n.d.). <https://www.luc.edu/its/aboutus/itspoliciesguidelines/hipaainformation/the18hipaaidentifiers/>

³ McCallister, E., Grance, T., & Scarfone, K. A. (2010). *Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)*. <https://doi.org/10.6028/nist.sp.800-122>

⁴ Dwork, C., & Roth, A. (2013). *The Algorithmic Foundations of Differential Privacy*. <https://doi.org/10.1561/9781601988195>

though differential privacy was a relatively new concept in 2010 (term was coined in 2006), researchers understood the particularly high costs that these privacy-preserving techniques would incur. Additionally, the authors may have related this back to fines or penalties an organization may face if a breach of PII or PHI data were to occur ranging from \$10,000-\$50,000 per violation and reaching \$5.6 billion per year⁵ in 2014.

Despite significant legal and regulatory developments since 2010, the authors' critique of the current definition of PII not being sufficient still holds true. Updates to U.S. privacy law, particularly HIPAA-related guidance from the Department of Health and Human Services (HHS), have expanded compliance expectations in some areas, but de-identification standards themselves have remained relatively static. The HIPAA Safe Harbor and Expert Determination methods continue to rely on the removal of a fixed set of 18 identifiers, even as re-identification techniques have become more sophisticated.⁶ The authors do make note of this stating, "...advances in the art and science of reidentification ...are rapidly rendering [existing de-identification techniques] obsolete."

The article is also notably ambiguous regarding how HIPAA applies (or doesn't) to de-identified data. Only in external research did it become clear to me that after removal of the 18 identifiers, the releasing party can no longer be held liable for inappropriate release of the remaining data⁶. Narayanan and Shmatikov seem to allude to this in *Lessons for Privacy Practitioners* but do not outright state that 'HIPAA Doesn't Apply to De-Identified Data.'

⁵ Adler, S. (n.d.). *Healthcare Data Breach Statistics*. The HIPAA Journal. <https://www.hipaajournal.com/healthcare-data-breach-statistics/>

⁶ Crawford, A., & McGraw, D. (n.d.). *Health Information Privacy after Dobbs*. American Bar Association. https://www.americanbar.org/groups/antitrust_law/resources/source/2024-oct/health-information-privacy-after-dobbs/

Sources

The 18 HIPAA identifiers | Information Technology Services (ITS): Loyola University Chicago. (n.d.-b).
<https://www.luc.edu/its/aboutus/itspoliciesguidelines/hipaainformation/the18hipaaidentifiers/>

Adler, S. (n.d.). *Healthcare Data Breach Statistics*. The HIPAA Journal. <https://www.hipaajournal.com/healthcare-data-breach-statistics/>

Crawford, A., & McGraw, D. (n.d.). *Health Information Privacy after dobbs*. American Bar Association.
https://www.americanbar.org/groups/antitrust_law/resources/source/2024-oct/health-information-privacy-after-dobbs/

Dwork, C., & Roth, A. (2013). *The Algorithmic Foundations of Differential Privacy*. <https://doi.org/10.1561/9781601988195>

Health Information Privacy after dobbs. (n.d.-a). https://www.americanbar.org/groups/antitrust_law/resources/source/2024-oct/health-information-privacy-after-dobbs/

McCallister, E., Grance, T., & Scarfone, K. A. (2010). *Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)*. <https://doi.org/10.6028/nist.sp.800-122>

Narayanan, A., & Shmatikov, V. (2010). Myths and fallacies of “personally identifiable information.” *Communications of the ACM*, 53(6), 24–26. <https://doi.org/10.1145/1743546.1743558>

Homework 3

Privacy Concerns of Amazon Alexa and the new Alexa+ product:

Voice assistants are systems that listen to and interpret natural language voice commands from users to perform a wide variety of tasks. Examples include Apple's Siri and most famously: Amazon's Alexa. By their nature, these devices are almost always "listening" to operate on-demand. This convenience has allowed Alexa to become the most widely used smart assistant in the world with over half a billion Alexa-enabled devices in 2023.

Since its invention, there have been significant privacy concerns surrounding Alexa's collection of data from millions of American households. Historically, researchers have been fast to point out the privacy flaws and vulnerabilities which make these devices harmful because of how they collect, process, access, and use personal data. These concerns are now being magnified by the 2026 rollout of Amazon's new **Alexa+** system, a generative AI-powered version of Alexa with even fewer user privacy controls.

Collection: Amazon Alexa/Echo devices have historically collected and used troves of user data. In 2022, PCMag and other researchers claimed Alexa captures the most data about a user compared to all other voice assistants at 34 data points. Alexa even captures this data about the contacts connected to your system or guests.

Consequences: All of the above data points and more go into building a virtual profile of users and their habits, which may in turn be used for advertising and profiling. In 2023, researchers from UC Davis published a collaborative study showing how this data collected from Alexa smart speakers was being used for ad-targeting. This report also notes that Amazon didn't mention this usage in their privacy policy until after publication.

Around the same time as this report, the Federal Trade Commission (FTC) charged Amazon with violating the Children's Online Privacy Protection Act (COPPA) by keeping children's voice recordings indefinitely and failing to honor parents' deletion requests. Amazon inevitably agreed to pay \$25 million as penalty for the allegations and were required to delete child accounts and voice recordings.

Conditions: Before 2025, Alexa included some privacy controls intended to limit data sent to Amazon's cloud. One such option, "Do Not Send Voice Recordings," allowed selected Echo models to process commands locally without transmitting recordings to Amazon's servers. Users could also choose "Don't Save Recordings," which kept data from being stored long-term. However, these controls were rarely used with fewer than 0.03% of users enabling local processing where available.

In 2019, a survey determined 41% of voice assistant users were concerned with the trust and privacy of the technology. In 2025, with the introduction of AI into this already intrusive tech, that number inflated to 73%. These percentages compared with the

dwindling number of users who enable privacy controls for the voice assistants, is a **classic display of a privacy paradox.**

Future: Starting in Spring 2026, the digital assistant is getting a major overhaul. Amazon is currently rolling out a system-wide upgrade called **Alexa+** to devices around the globe. This mandatory upgrade aims to increase the voice assistant's capabilities by implementing more conversational AI into the platform.

Immediately after the beta test rollout, users noticed that the option to turn off data processing has been completely removed. Starting in 2026, Amazon will be removing the "Do Not Send Voice Recordings" option, which means all recorded voice commands will be automatically sent to Amazon for processing and analysis. The company is also changing how the "Do Not Save Voice Recordings" option works. Because the new Alexa+ relies on generative AI capabilities hosted in the cloud, Amazon has stated that local processing is no longer supported. Users are forced to choose between sharing all voice data with Amazon or disabling key device functionality.

While Amazon states that voice requests are encrypted and that user privacy remains a priority, the company's historical use of interaction data for ad-targeting without clear upfront disclosure has raised concerns about transparency in this new platform. Also, as described in Narayanan & Shmatikov, deanonymization of data alone is not enough either.

The future of this technology remains unknown; however, it is likely that an additional use case of consumer voice data will be helping Amazon train its AI systems. Amazon's Trust & Privacy page seems to focus more on the features of Alexa+ instead of the implications for the use. This may deter many users from controlling their data.

Analysis: The evolution from Alexa to Alexa+ illustrates several critical privacy concerns. First, this forced transition demonstrates how privacy is increasingly treated as optional by the controlling businesses rather than a fundamental user right. This upgrade is still in its infancy, but the elimination of privacy-protective options is an obvious erosion of what few controls the consumer previously had.

Second, this case highlights the failures of consent-based privacy frameworks when dealing with shared spaces like the home. Voice assistants collect audio data from all parties and visitors, not merely the account holder who consented to data collection. Research examining multi-user privacy dynamics in smart homes has documented how voice assistants create "bystander privacy" concerns, wherein individuals who did not consent to data collection are nonetheless subject to surveillance (*Geeng & Roesner*).

Finally, the shift toward mandatory cloud processing reflects a trend that prioritizes corporate data access over user privacy. This feature may marginally increase functionality while greatly removing data from the user's control. While Amazon frames

this requirement as necessary for generative AI capabilities, alternative approaches could preserve privacy while still enabling advanced functionality.

Sources

Because Alexa+ is so new, official scholarly sources on its data usage is limited.

Alexa just became less private — what the March 28 Amazon changes mean. (n.d.-a).

<https://www.forbes.com/sites/daveywinder/2025/03/28/alexa-just-became-less-private-what-the-march-28-amazon-changes-mean/>

Amazon's Alexa AI upgrade remains in early access. CNET survey shows 73% of device Users Express Privacy Worries. CNET. (n.d.-a).

<https://www.cnet.com/home/smart-home/amazons-alexa-ai-upgrade-remains-in-early-access-cnet-survey-shows-73-of-device-users-express-privacy-worries/>

Amazon's privacy ultimatum starts today: Let echo devices process your data or stop using Alexa. CNET. (n.d.-b).

<https://www.cnet.com/home/security/a-privacy-ultimatum-starts-today-let-amazon-echo-process-your-data-or-stop-using-it/>

Bonifield, S. (2026, January 12). *Amazon has started automatically upgrading prime members to Alexa plus.* The Verge.

<https://www.theverge.com/news/860581/amazon-prime-alexa-plus-automatic-upgrade>

Chung, H., Iorga, M., Voas, J., & Lee, S. (2017). "Alexa, can I trust you?" *Computer*, 50(9), 100–104.

<https://doi.org/10.1109/mc.2017.3571053>

Cohen, J. (2022, March 30). *Amazon's Alexa collects more of your data than any other Smart assistant.* PCMag.

<https://www.pcmag.com/news/amazons-alexa-collects-more-of-your-data-than-any-other-smart-assistant>

Farrell, N. (2026, January 27). *Amazon is rolling out alexa+ to all users. but not everyone wants it.* Wired.

<https://www.wired.com/story/alexa-plus-early-access-rollout-2026/>

Geeng, C., & Roesner, F. (2019). Who's in control? *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. <https://doi.org/10.1145/3290605.3300498>

Hernández Acosta, L., & Reinhardt, D. (2024). "Alexa, How Do You Protect My Privacy?" *A Quantitative Study of User Preferences and Requirements about Smart Speaker Privacy Settings.* <https://doi.org/10.2139/ssrn.4990104>

Iqbal, U., Bahrami, P. N., Trimananda, R., Cui, H., Gamero-Garrido, A., Dubois, D. J., Choffnes, D., Markopoulou, A., Roesner, F., & Shafiq, Z. (2023). Tracking, profiling, and ad targeting in the alexa echo smart speaker ecosystem. *Proceedings of the 2023 ACM on Internet Measurement Conference*, 569–583. <https://doi.org/10.1145/3618257.3624803>

Nguyen, S. T., Jillson, E., Saqib, Fyi, & Couillaud, P. (2025, April 1). *Hey, Alexa! what are you doing with my data?* Federal Trade Commission. <https://www.ftc.gov/business-guidance/blog/2023/06/hey-alexa-what-are-you-doing-my-data>

Press, T. A. (2025, March 19). *Amazon ends little-used privacy feature that let echo users opt out of sending recordings to company.* AP News. <https://apnews.com/article/amazon-privacy-echo-7fb3c19fa7f64bde5c5be259f8b23ee>

Scharon Harding, A. T. (2025, March 17). *Everything you say to your echo will soon be sent to Amazon, and you can't opt out.* Wired. <https://www.wired.com/story/everything-you-say-to-your-echo-will-be-sent-to-amazon-starting-march-28/>

Shibu, S. (2026, January 30). *Amazon Echo ends "do not send Voice Recordings" option.* Entrepreneur.

<https://www.entrepreneur.com/business-news/amazon-echo-ends-do-not-send-voice-recordings-option/488677>

Statement of commissioner Alvaro M. Bedoya joined by ... (n.d.-b). https://www.ftc.gov/system/files/ftc_gov/pdf/Bedoya-Statement-on-Alexa-Joined-by-LK-and-RKS-Final-1233pm.pdf

Study shows Alexa Invades Privacy, collects user data for ad-targeting. (n.d.-c). <https://engineering.ucdavis.edu/news/study-shows-alexa-invades-privacy-collects-user-data-ad-targeting>