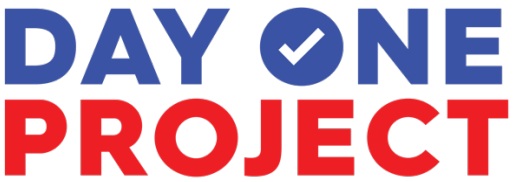
Have Your Data and Use It Too: A Federal Initiative for Protecting Privacy while Advancing AI

Roxanne Heston and Helen Toner

January 23, 2020



dayoneproject.org

Summary

The next administration should aim to make the United States a world leader in privacy-preserving machine learning (PPML), a collection of new artificial intelligence (AI) techniques capable of providing the benefits of machine learning while minimizing data-privacy concerns.

By some estimates, improvements to the speed, accuracy, and scale of AI could augment global GDP by 14%, or $15.7 trillion, by 2030.[1] Yet Americans fear that expansion of AI will have moderate to severe negative consequences.[2] They are particularly concerned about the privacy implications of how companies and agencies use personal data to generate new developments.

To assuage these concerns, we recommend targeted initiatives aimed at bringing PPML techniques to maturity:

(1) **Invest in PPML research and development (R&D).** The next administration should issue a Presidential Memorandum on Day One making PPML a priority and establishing a goal of making PPML use a simple, default option for all applications. The next administration should support funding federal agencies (such as the National Science Foundation (NSF) and the Defense Advanced Research Projects Agency (DARPA)) to invest in R&D efforts that will make PPML techniques applicable to a wide variety of industries and users.

(2) **Identify compelling opportunities to apply PPML techniques at the federal level.** U.S. researchers are making remarkable progress on AI capabilities. The U.S. government must ensure it matches their pace of progress in the realm of AI data security. On Day One, the next term should commission reports on the potential for PPML to improve public services provided by the federal government involving sensitive citizen data. Bodies that could write such reports include the Department of Defense (DOD)'s Joint Artificial Intelligence Center (JAIC), the Congressional Research Service (CRS), the Government Accountability Office (GAO), and the National Academies. The Federal Trade Commission, as well as other federal banking and financial-regulation agencies, should also create resources to encourage private-sector PPML adoption.

(3) **Create frameworks and technical standards to facilitate wider deployment of PPML techniques.** The next administration should task the National Institute of

[1] PWC, "AI to drive GDP gains of $15.7 trillion with productivity, personalisation improvements", June 27, 2017, https://press.pwc.com/News-releases/ai-to-drive-gdp-gains-of--15.7-trillion-with-productivity--personalisation-improvements/s/3cc702e4-9cac-4a17-85b9-71769fba82a6.

[2] Baobao Zhang and Allan Dafoe, "Chapter 2: General attitudes towards AI", *Artificial Intelligence: American Attitudes and Trends*, Center for the Governance of AI, Future of Humanity Institute, University of Oxford, https://governanceai.github.io/US-Public-Opinion-Report-Jan-2019/general-attitudes-toward-ai.html.

Standards and Technology (NIST) with developing a framework to evaluate existing PPML measures and develop guidance documents for use of PPML techniques in the public sector. NIST should also collaborate with external stakeholders to develop technical standards that support the broad application of PPML techniques. The next president should request that NIST receive dedicated funding to carry out this work and should also direct individual agencies to support development of PPML use guidelines tailored to their circumstances.

1. Challenge

1.1 Data sharing and consumer privacy

Many routine activities—such as accessing personal finances or receiving citizenship benefits—require individuals to share personally identifying information (PII) like credit-card details, Social Security numbers, physical addresses, and medical records. Sharing these details is like being required to periodically share copies of your house keys with strangers: it is hard to track where they go after you give them out, you have to trust the recipient, and you expose your most important assets to theft. But you cannot change your PII like you can change your locks. And once thieves have your PII, they can steal from you from anywhere in the world.

Another, newer source of sensitive data relates to consumer use of digital services. Digital messaging services log user communications, digital vendors log customer purchasing behavior, and search engines log user queries. User data can be used to build detailed profiles of—and better tailor products and services to—individual users. But data and profiles are also sometimes sold to third-party companies that use this information in ways that adversely affect consumers. For instance, predatory lenders can target particularly vulnerable people and healthcare companies can avoid accepting clients with preconditions. In the worst cases, bad actors can illegally access and exploit user data and profiles for blackmail and theft. But just as with PII, consumers have limited control over who has access to their use information or how that information is used.

Thanks to widespread sharing of sensitive data, identity theft has become a significant problem in the United States. The vast majority of identity theft involves a nefarious actor misusing (or attempting to misuse) an existing account like a credit card or bank account.[3] 23% of American internet users reported having experienced one or more instances of identity theft as of October 2018.[4] The Department of Justice estimates that identity

[3] Erika Harrell, "Victims of Identity Theft, 2016", NCJ 251147, Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice (January 2019).
[4] Statista, "Share of internet users in the United States who have been victim of online identity theft as of October 2018", https://www.statista.com/statistics/763130/internet-identity-theft-usa/.

theft cost its victims an estimated $15.4 billion in 2014.[5] Insufficient data-privacy protections threaten companies as well as consumers. Privacy concerns prompt consumers to seek out applications that promise greater security and to leave those that are lacking.[6] Data sharing also introduces the risk of faulty security measures, letting adversaries exploit the companies' clientele and operational details.[7]

1.2 Shortcomings of existing approaches to data privacy

Conventional ways of improving data privacy are costly and leave much to be desired. An MIT study found that researchers could identify 95% of people in a dataset comprised solely of mobile and transit logs, even when those data were classically anonymized.[8] The European Union's General Data Protection Regulation (GDPR) attempts to protect consumers by requiring companies to make collection, use, and deletion of personal data user-accessible. Attempting to accommodate these demands in Europe cost U.S. multinational companies billions of dollars. As a regulation, not a technical fix, the GDPR simply makes personal-data breaches costly to companies and transparent to consumers rather than providing means to avoid breaches outright.[9,10]

Implementing more stringent data-sharing restrictions is not the solution. The fact that companies spend tens of billions of dollars annually on data acquisition and analysis indicates that access to "big data" is enormously valuable for the private sector.[11] Consumers in turn benefit from data-driven insights and product improvements. Drastically limiting data sharing would drastically limit the potential benefits of big data

---

[5] Erika Harrell, "Victims of Identity Theft, 2014", NCJ 248991, Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice (September 2015).

[6] A Pew Research Center found that 79% of adult respondents say that are not too or not at all confident that companies will admit mistakes and take responsibility if they misuse or compromise personal information. Yet only 19% of those Americans feel they can control the data companies collect about them. Accordingly, privacy concerns caused Facebook to lose nearly half of its surveyed quitters in 2013 and, conversely, doubled downloads of the encrypted messaging app Signal in 2017. Sources: Brooke Auxier, *et al.*, "Americans and Privacy: Concerned, Confused, and Feeling Lack of Control Over Their Personal Information", Pew Research Center, November 15, 2019, https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/; Stefan Steiger, *et al.*, "Who Commits Virtual Identity Suicide? Differences in Privacy Concerns, Internet Addition, and Personality Between Facebook Users and Quitters", *Cyberpsychology, Behavior, and Social Networking* 16, no. 9 (September 2013); Rani Molla, "Trump fears have helped double downloads for Signal, the private messaging app", *Vox*, April 4, 2017, https://www.vox.com/2017/4/4/15124316/americans-privacy-signal-downloads-chart.

[7] To understand how risk increases as data sharing increases, see National Institute of Standards and Technology , "Framework for Improving Critical Infrastructure Cybersecurity", U.S. Department of Commerce, April 16, 2018, https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf.

[8] Rob Matheson, "The privacy risks of compiling mobility data", *MIT News*, December 7, 2018, https://news.mit.edu/2018/privacy-risks-mobility-data-1207.

[9] Mehreen Khan, "Companies face high cost to meet new EU data protection rules", *Financial Times*, November 19, 2017, https://www.ft.com/content/0d47ffe4-ccb6-11e7-b781-794ce08b24dc.

[10] Bernard Marr, "GDPR: The Biggest Data Breaches And The Shocking Fines (That Would Have Been)", *Forbes*, June 11, 2018, https://www.forbes.com/sites/bernardmarr/2018/06/11/gdpr-the-biggest-data-breaches-and-the-shocking-fines-that-would-have-been/#7509bd366c10.

[11] Business Wire, "Double-Digit Growth Forecast for the Worldwide Big Data and Business Analytics Market Through 2020 Led by Banking and Manufacturing Investments, According to IDC", October 3, 2016, https://www.businesswire.com/news/home/20161003005030/en/Double-Digit-Growth-Forecast-Worldwide-Big-Data-Business.

and the AI techniques that big data support. Researchers anticipate that replicating the GDPR in the United States would have significant negative effects on AI innovation in industry.[12] The preferable alternative is to ensure that data is collected, stored, and shared in ways that truly protect user privacy.

2. Opportunity

Although the increasing use of AI is seen as a threat to data privacy, it could actually be a solution. Emerging methods in machine learning (the prevailing approach to AI development today) make it possible for data to be used in valuable ways without compromising privacy (see Appendix for technical details). These methods could eliminate the need for companies to access PII or view user data in order to process transactions or improve services. Privacy-preserving machine learning (PPML) techniques can satisfy some of the data needs of organizations while addressing privacy concerns of the public—thereby offering a more attractive and effective approach than policies like the GDPR.

Federal agencies are uniquely situated to advance PPML techniques. They can coordinate relevant research efforts, standardize implementation, and prompt and enforce adoption. Federal investments in PPML are likely to yield a multitude of valuable applications—such as improving recommender systems and increasing product and service transparency—that could be useful not just to the private sector and its customers but also to government bodies, researchers, and other countries.

3. Proposed action

The next administration should seize the opportunity to promote PPML development and implementation now, before machine learning becomes even more widely implemented. Doing so effectively will require integrating PPML into AI systems at the outset, rather than simply attempting retroactive vulnerability patches.

PPML techniques, like machine learning broadly, have a wide range of potential use cases. To address the country's foremost concerns, the next administration should focus on applications that make new uses of AI possible without triggering privacy concerns, give U.S. companies a competitive edge over their foreign counterparts, and/or reduce cybersecurity risks by protecting user data while preserving its usefulness.

[12] Nick Wallace and Daniel Castro, "The Impact of the EU's New Data Protection Regulation on AI", Center for Data Innovation (March 2018).

To achieve these goals, the U.S. government should:

(1) **Invest in PPML research and development (R&D)** to make PPML use a simple, default option for all applications;
(2) **Identify compelling opportunities to apply PPML techniques at the federal level,** including by seeking out demonstration applications that can be implemented using government data; and
(3) **Create frameworks and technical standards to facilitate wider deployment of PPML techniques** and encourage the consideration of the systems requirements for doing so.

Additional detail on each of these activities is provided below.

3.1 Invest in PPML research and development (R&D)

PPML techniques started are in the early days of adoption. Though certain components of PPML have existed (at least conceptually) since the late 1970s,[13] PPML only started receiving meaningful attention around 2009. Some PPML techniques, like federated learning, only appeared as late as 2016.[14]

Continued progress in PPML demonstrates the organic interest of researchers in developing this topic area. Through federal funding and messaging, the U.S. government can tap that interest and spur additional progress in PPML techniques. Ambitious government investment would be particularly valuable given the nascent stage of PPML. The eventual goal would be simple and default application of PPML techniques in all relevant domains, much as encryption is a simple and default practice in web browsing today. Accordingly, we recommend that federal funding for PPML R&D prioritize projects that make PPML techniques more generally applicable.

The NSF's Directorate for Computer and Information Science and Engineering (CISE) is one fitting place to house PPML R&D efforts, given the directorate's goals of (1) promoting understanding of the principles and uses of advanced computing and (2) contributing to transparent participation in an information-based society. CISE—and related NSF directorates and programs—can support advancement of PPML through grants and proposal competitions. Indeed, PPML R&D fits naturally into some existing NSF programs. For instance, NSF programs on Cyber-Physical Systems and the Secure

[13] "Secure multi-party computation", *Wikipedia*, last modified December 6, 2019, https://en.wikipedia.org/wiki/Secure_multi-party_computation.
[14] To track the evolution of PPML research, we queried arXiv (https://arxiv.org/)—an online repository for scientific papers widely used among those working in AI—with the phrases "Privacy Preserving Machine Learning", "Differential Privacy", and "Federated Learning".

and Trustworthy Cyberspace Frontiers both seek innovative proposals for fundamental cybersecurity research. [15] By obscuring or not sharing data used for AI training, PPML makes data less vulnerable to confidentiality breaches in confidentiality—thereby improving cybersecurity. NSF should consider explicitly mentioning PPML in solicitations for cybersecurity R&D proposals.

PPML R&D also aligns with efforts at DARPA. DARPA's AI Next Campaign seeks to invest in basic and applied research that will create "new, game-changing AI technologies for U.S. national Security." With its explicit emphasis on robustness and invulnerability, the AI Next Campaign could place specific attention on cultivating PPML techniques. The AI Exploration component of the Campaign could even surface entirely new approaches to PPML, mirroring DARPA's previous funding for research into multiparty-computation techniques.[16]

Finally, government-sponsored prize competitions can be an especially effective way to orient private-sector researchers toward certain priority topics—both by offering a financial incentive for investment and by increasing the visibility of messaging on those topics. The 21st Century Grand Challenge series,[17] a former project of the White House Office of Science and Technology Policy (OSTP), spurred nationwide investments in early-stage, security-relevant fields. Federal agencies worked with OSTP to catalyze brain research, solar energy, and asteroid identification, prompting non-governmental entities to follow their leads.[18] The next administration should leverage the unique funding-messaging combination of prize competitions to accelerate PPML R&D.

3.2 Identify compelling opportunities to apply PPML techniques at the federal level

The U.S. government is just starting to determine where AI can and should be applied. This determination is often carried out on an ad-hoc basis by those federal employees with AI knowledge in their respective parts of government. Recent initiatives like the DOD's Joint Artificial Intelligence Center (JAIC) have been established to more systematically evaluate where and how AI could be most usefully developed and deployed, and to ensure that AI technology is used securely and to the benefit of the American public.[19] The next administration should extend these initiatives to the many

[15] "Secure and Trustworthy Cyberspace Frontiers (SaTC Frontiers)", National Science Foundation, n.d., https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505641&org=CISE&from=home.
[16] Defense Advanced Research Project Agency, "AI Next Campaign", U.S. Department of Defense, n.d., https://www.darpa.mil/work-with-us/ai-next-campaign.
[17] "21st Century Grand Challenges", White House Office of Science and Technology Policy, n.d., https://obamawhitehouse.archives.gov/administration/eop/ostp/grand-challenges.
[18] *Ibid.*
[19] The JAIC includes "defend[ing] U.S. critical infrastructure from malicious cyber activity" and "leading in military AI ethics and safety" in its holistic approach to accelerating the delivery and adoption of AI. Similarly, the DOD's Defense Innovation Board includes "assess[ing] cyber security vulnerabilities of advanced weapons" in its recommendations. Sources: Joint Artificial

other agencies that intend to employ AI or are tasked with responsible use. Indeed, the next administration should position the U.S. government to lead by example when it comes to PPML, ensuring that progress in AI data security matches progress in AI capabilities.

To this end, we recommend that the next administration commission one or more reports on the potential for PPML to improve public services provided by the federal government involving sensitive citizen data. Bodies that could write such reports include the JAIC, CRS[20] and GAO[21] (for legislators), and the National Academies[22] (for broader internal dissemination). In addition, the Government Effectiveness Advanced Research (GEAR) Center could issue a request for information to source PPML application suggestions from the private sector.

Some possible applications are already visible. The U.S. Census Bureau published its own differentially private dataset in 2016[23] and is conducting the 2020 Census using PPML techniques.[24] The Bureau's xD project seeks to apply AI solutions to government services, and hence could reasonably add PPML research and applications to its portfolio. For financial products and services, the Federal Trade Commission (FTC)—per its mandate to protect consumer financial privacy[25]—could recommend or require the use of PPML techniques by companies that employ machine learning. The FTC, along with other agencies involved in banking and financial regulation, could update model privacy forms to indicate the use of privacy-preserving practices where applicable. [26]

3.3 Create frameworks and technical standards to facilitate wider deployment of PPML techniques

Much of the value of PPML can only be realized if consumers understand and trust the protections in place, even if only at a high level. The lock displayed to the left of URLs on secure web pages is a well-known symbol that demonstrates compliance with a

---

Intelligence Center, "Vision: Transform the DoD Through Artificial Intelligence", U.S. Department of Defense, n.d., https://dodcio.defense.gov/About-DoD-CIO/Organization/JAIC/; Defense Innovation Board, "Recommendations", U.S. Department of Defense, n.d., https://innovation.defense.gov/Recommendations.aspx.

[20] Under its Resources, Science and Industry division.

[21] Under its Technology & Science portfolio.

[22] Under its Division on Engineering and Physical Sciences.

[23] Erica Portnoy, Gennie Gebhart, and Starchy Grant, "Facial Recognition, Differential Privacy, and Trade-Offs in Apple's Latest OS Releases", Electronic Frontier Foundation, September 27, 2016, https://www.eff.org/deeplinks/2016/09/facial-recognition-differential-privacy-and-trade-offs-apples-latest-os-releases/.

[24] John M. Abowd, "Why the Census Bureau Adopted Differential Privacy for the 2020 Census of Population", Harvard University Privacy Tools Project, 2014, https://privacytools.seas.harvard.edu/why-census-bureau-adopted-differential-privacy-2020-census-population.

[25] Federal Trade Commission, "Financial Privacy: Protecting Consumers' Financial Privacy", n.d., https://www.ftc.gov/news-events/media-resources/protecting-consumer-privacy/financial-privacy.

[26] For an example, see the model privacy form available at https://www.ftc.gov/system/files/documents/rules/privacy-consumer-financial-information-financial-privacy-rule/privacymodelform_optout.pdf.

specific standard in web security. A similar system could exist in AI. With agreed-upon standards for PPML implementation, developers could highlight and build consumer trust in machine-learning systems that use personal data responsibly. The Department of Commerce's National Institute of Standards and Technology (NIST) is well positioned to lead development of widely used and accepted tests, benchmarks, and standards.

NIST currently conducts research to determine "how to measure and enhance the security and trustworthiness of AI systems."[27] To that end, NIST issued a request for information in May 2019 regarding federal efforts to develop standards for AI technologies.[28] NIST's AI research program is part of a broader effort to "issue a plan for Federal engagement in the development of technical standards and related tools in support of reliable, robust, and trustworthy systems that use AI technologies."[29] This program also mirrors the work NIST is doing at the international level: namely, working with the International Organization for Standardization (ISO)[30] to "ensure innovation, public trust and confidence in systems that use AI technologies."[31]

NIST is well positioned to assess PPML technique implementation throughout the federal government. In particular, NIST could develop a framework to evaluate how well existing machine-learning systems in the federal government incorporate PPML techniques. Individual agencies could receive funding to generate tailored guidelines for improving the operations of these systems to meet data-privacy standards established by NIST. Relatedly, NIST's National Cybersecurity Center of Excellence could generate Cybersecurity Practice Guides designed to support public-sector use of PPML techniques.

Despite the importance of NIST's priorities, President Trump's FY2020 budget request proposed a nearly one-third cut to the agency's $1 billion budget. The proposed cuts included a 16% reduction to NIST's Scientific and Technical Research and Services budget area and a near halving of the area's Standards Coordination and Special Programs account.[32] This funding facilitates, respectively, activities like a differential

---

[27] National Institute of Standards and Technology, "Artificial Intelligence", U.S. Department of Commerce, November 18, 2019, https://www.nist.gov/topics/artificial-intelligence.

[28] National Institute of Standards and Technology, "Request for Information about Federal Engagement in Artificial Intelligence Standards", U.S. Department of Commerce, n.d., https://www.nist.gov/topics/artificial-intelligence/request-information-about-federal-engagement-artificial-intelligence.

[29] "Executive Order 13859 of February 11, 2019, Maintaining American Leadership in Artificial Intelligence", *Code of Federal Regulations*, title 3 (2019): 3967–3972, https://www.govinfo.gov/content/pkg/FR-2019-02-14/pdf/2019-02544.pdf.

[30] International Organization for Standardization, "ISO/IEC JTC 1/SC 42: Artificial intelligence", n.d., https://www.iso.org/committee/6794475.html.

[31] National Institute of Standards and Technology, "Artificial Intelligence".

[32] Adria Schwarber, "FY20 Budget Request: National Institute of Standards and Technology", America Institute of Physics, April 3, 2019, https://www.aip.org/fyi/2019/fy20-budget-request-national-institute-standards-and-technology.

privacy challenge[33] and an international (ISO) partnership for AI standards. Fortunately, Congress actually *increased* NIST spending in its FY2020 appropriations.[34] To allow NIST to successfully carry out its existing and proposed AI activities, the next president should submit a next budget request that matches or increases FY2020 appropriated funds for NIST efforts related to AI standards.

4. Conclusion

The next administration can vastly influence the trajectory of how AI affects user privacy. The United States has both the ability to develop cutting-edge AI technologies and the responsibility to ensure their responsible use. AI is arguably one of today's most important emerging technologies, and, according to Americans, privacy is the most important area within AI to get right. To ensure the country is at the forefront of responsible AI, the next administration should integrate, promote, and standardize the development and adoption of privacy-preserving machine learning techniques.

[33] Communications Technology Laboratory, "2018 Differential Privacy Synthetic Data Challenge", Public Safety Communications Research Division, National Institute of Standards and Technology, U.S. Department of Commerce, December 5, 2019, https://www.nist.gov/communications-technology-laboratory/pscr/funding-opportunities/prizes-challenges/2018-differential.
[34] "Spending Deal Buoys Science Agency Budgets", America Institute of Physics, December 17, 2019, https://www.aip.org/fyi/2019/spending-deal-buoys-science-agency-budgets.

A. How PPML works

A.1 Machine learning

The most notable developments in AI are coming from a category of approaches called machine learning, in which programmers set computers up to learn how to do new tasks instead of manually writing out the steps they should perform. To fuel this process, machine-learning researchers develop some base code (an "algorithm"), put that code on a powerful computer (one with plenty of "compute"), and then feed the algorithm lots of relevant examples from which to learn ("training data"). The trained algorithm, called a "model," is what AI-based systems use to operate. A model is often more adept the more training data it is based on.

In just the past five years, machine learning has allowed computers to perform tasks previously thought exclusive to human brains. A current hallmark of this era of investment is the 2015 achievement of world-class performance at the game Go,[35] although many promising areas of research since have attempted feats outside the gaming realm. [36] *Machine Learning for Policymakers* (Buchanan and Miller 2017) provides a more detailed explanation of how machine learning works.[37]

Since researchers try to teach computers how to do what humans can do, the most relevant training data are often pieces of information generated or used by people. For instance, AI use audio clips and photos to become adept at speech and image recognition; patient files to make medical diagnoses; chat logs to generate chatbot text; and user content and behavior to optimize newsfeeds or advertisements. When it comes to machine learning, researchers only need access to training data to extract lessons from the points in aggregate; they do not need to be able to view individual data points.

It was long impossible to decouple learning-relevant data access to data access in its entirety. But over the past several years a collection of privacy-preserving machine learning (PPML) techniques has developed that allows data to be used without being revealed, permitting the best of both worlds.

A.2 Privacy-preserving machine learning

"Privacy-preserving machine learning" (PPML) refers to a class of machine-learning approaches, frameworks, and techniques that are designed to achieve similar

---

[35] Sam Byford, "Why is Google's Go win such a big deal?", The Verge, March 9, 2016, https://www.theverge.com/2016/3/9/11185030/google-deepmind-alphago-go-artificial-intelligence-impact.

[36] Javier Couto, "The major advancements in Deep Learning in 2018", Tryo Labs, December 19, 2018, https://tryolabs.com/blog/2018/12/19/major-advancements-deep-learning-2018/.

[37] Ben Buchanan and Taylor Miller, *Machine Learning for Policymakers: What It Is and Why It Matters*, Belfer Center for Science and International Affairs, Harvard Kennedy School, Harvard University (June 2017).

performance to regular machine-learning algorithms while avoiding many of the associated privacy challenges.

Two of the most promising types of PPML are federated learning (a form of multiparty computation) and differential privacy. Brief explanations of each follow. The research group Facebook AI offers information about these and other PPML techniques in their free Udacity course, "Secure and Private AI."

Federated learning

Common practice in machine learning is to gather training data in a centralized database and then feed that data to an algorithm for analysis. Federated learning is an emerging alternative. This distributed machine-learning technique allows devices to keep data private by training a machine-learning algorithm on the device holding the data instead of sending the data to a company to be trained on that company's servers.

In 2017, Google developed a federated-learning algorithm for its Android phone keyboard. Following a conventional machine-learning pathway, the model making predictive-text recommendations for the phone would sit in a cloud server, import information about user behavior (*i.e.*, whether or not users accepted its predictive-text suggestions), and then learn from the behavior data to develop a new-and-improved version of the model. With federated learning, the current version of the model downloads to a user's phone, learns from the specific user's behavior, and then sends lessons back to the centralized model where those lessons are integrated with lessons from other phones. In other words, the predictive-text algorithm still uses user behavior to improve but avoids the need for user data to ever be gathered and stored in a central location where a human could access the data.

Federated learning falls into a broad category of related approaches collectively called secure multiparty computation (MPC). MPC techniques make it possible for multiple parties to cooperate in developing an AI system without exchanging privately held information.[38] Some MPC differs from federated learning in that the information flow is reciprocal. Rather than drawing conclusions on one side and updating a centralized model, MPC can be entirely decentralized in a way that allows participants to draw conclusions together by sharing only relationships between data. The first notable use case of MPC, a 2008 sugar-beet auction in Denmark, demonstrated the possibility of using MPC in higher-stakes scenarios like secure voting systems.[39] MPC has also

[38] Ben Garfinkel, "Recent Developments in Cryptography and Possible Long-Run Consequences," *forthcoming*.
[39] Peter Bogetoft, *et al.*, "Secure Multiparty Computation Goes Live", International Association for Cryptologic Research (2008) https://eprint.iacr.org/2008/068.pdf.

facilitated collaboration among parts of government with private datasets, such as between branches of the Estonian government in 2015. [40] MPC has the potential to permit the inter-agency use of sensitive compartmentalized information within the United States.[41]

Differential privacy

Differential privacy shares training data just as data are shared in conventional machine-learning approaches—but differential privacy makes those data indecipherable to humans. By strategically injecting small amounts of randomness into training data, researchers can make it impossible to draw conclusions about any particular individual while preserving the ability to draw useful conclusions from the dataset as a whole. Apple, known for its comparatively strong security practices, employs differential privacy techniques to learn about the experiences of iPhone users.[42]

[40] "The European Commission highly evaluates the project on privacy-preserving computation", Cybernetica, October 9, 2015, https://cyber.ee/news/2015/10-09/.
[41] Yehua Lendell and Benny Pinkas, "Secure multiparty computation for privacy-preserving data mining", *Journal of Privacy and Confidentiality* 1, no. 1 (2009).
[42] "Learning with Privacy at Scale", Machine Learning Journal, Apple Inc., https://machinelearning.apple.com/2017/12/06/learning-with-privacy-at-scale.html.

About the authors

Roxanne Heston is a Research Analyst at Georgetown's Center for Security and Emerging Technology (CSET). She is concurrently a master's student in Georgetown's Security Studies Program. Roxanne previously assisted the research of former SECNAV Richard Danzig, Dr. Ben Buchanan, and Oxford University's Center for the Governance of AI. She received a B.S. in Economics with honors on a full scholarship from Tulane University, where she was an Altman Scholar in International Studies & Business.

Helen Toner is Director of Strategy at Georgetown's Center for Security and Emerging Technology (CSET). She previously worked as a Senior Research Analyst at the Open Philanthropy Project, where she advised policymakers and grantmakers on AI policy and strategy. Between working at Open Philanthropy and joining CSET, Helen lived in Beijing for nine months, studying the Chinese AI ecosystem as a Research Affiliate of Oxford University's Center for the Governance of AI.

About the Day One Project

The Day One Project is dedicated to democratizing the policymaking process by working with new and expert voices across the science and technology community, helping to develop actionable policies that can improve the lives of all Americans, and readying them for Day One of a future presidential term. For more about the Day One Project, visit dayoneproject.org.