Privacy Is Power

In 1992, a group of pioneering policymakers and technologists gathered at an international conference to chart the course of a distributed computing network. The arrival of digital platforms that would enable the near-instantaneous sharing of information by millions of people and institutions globally was just around the corner. The technology seemed poised to change the world, these visionaries understood, yet the direction of change was far from certain. With an eye on the future, they decided to give history a push. They asked critical questions about the technology's maturity and forged partnerships with governments, industry sectors, and academia to steer the development of the Internet in ways that enshrined democratic values.

If this anecdote sounds too good to be true, that's because it is. In reality, few of the policymakers who were present at the creation of the Internet predicted that the hypertext transfer protocol used to load webpages would prove dominant, and even fewer considered what it might take to govern the Internet at scale. Present-day Web users are living with the consequences of their inaction: weaponized social media, cyber-intrusions that prey on the vulnerabilities of Internet architecture, the buying and selling of informed predictions about individual Internet users' future behavior, and information monopolies that threaten democratic discourse online.

History is rarely forgiving, but as we adopt the next phase of digital tools, policymakers can avoid the errors of the past. Privacy-enhancing technologies, or PETs, are a collection of technologies with applications ranging from improved medical diagnostics to secure voting systems and messaging platforms. PETs allow researchers to harness big data to solve problems affecting billions of people while also protecting privacy. As their use becomes widespread, a new paradigm will emerge, in which private data is more readily available for research and problem solving and public data on private citizens is better protected. Much like the Internet itself, PETs have become a dividing line between democratic and authoritarian governments with implications for privacy and accountability. With foresight, however, the United States and its allies can realize the benefits of PETs while preventing the most dangerous outcomes. Indeed, the announcement at U.S. President Joe Biden's recent Summit for Democracy of a joint effort by the United Kingdom and the United States to create a grand challenge aimed at fostering innovation in PETs signals the importance some advanced democracies are attaching to this effort.

PROTECTING ONLINE PRIVACY

PETs are a class of techniques, methods, and mathematical approaches that allow one person to answer a question using data owned by another individual (or organization), without the asker learning anything more than the answer to the question—thus protecting sensitive information used in the process. Their emergence in recent years has upended the conventional wisdom that the goals of data analysis and privacy are at odds with each other, because now data owners have to decide only whether to allow someone to answer a specific question, as opposed to deciding whether to reveal their data.

Consider the researcher trying to figure out whether Americans are losing jobs to robots. Economists have long struggled to determine whether automation is a major driver of unemployment, in part because of a bureaucratic partition: the U.S. Bureau of Labor Statistics collects information on which companies are buying robots, and the U.S. Census Bureau collects information on which people are losing jobs, but these data sets are so sensitive that the two agencies will not share them with each other, let alone with private researchers. PETs, however, allow an analyst to leverage data from both agencies, without requiring those agencies to send a copy of their data to anyone. That is to say, these agencies can collectively empower an economist to answer an important question without compromising privacy because the data remains undisclosed.

How does the technology work? One approach is called "federated learning." This machine-learning technique trains algorithms on individual devices instead of centralized databases, which means that user data isn't stored or readily accessible in a central location. Another approach is called "differential privacy." The idea there is to use randomness to obscure personal information in large data sets. By adding noise to the data, researchers can give plausible deniability to any particular individual in the data set and prevent the reverse engineering of research results to identify individual data. Companies and governments are piloting applications of both techniques. Federated learning allows Apple to use insights gleaned from the experiences of millions of iPhone owners to improve its voice recognition and predictive-typing capabilities, without collecting or storing any of the owners' private data in its central servers. In other words, your iPhone knows that when you type "look at what bisen said," you usually mean "look at what Biden said," because millions of other users have made that correction, not because your personal messages are being collected and examined.

In 2016, after researchers at the U.S. Census Bureau learned that commercially available data could be combined with census outputs to identify respondents, the bureau used differential privacy to protect some of the 2020 census data it released. Examining published census data protected by a differential privacy-based system is a bit like viewing a pointillist painting: as long as you're standing back, you can see the bigger picture that the dots combine to depict, but if you try to peer too closely, the technologies automatically blur the individual dots.

Another PET tool is secure multiparty computation, a cryptographic protocol that enables researchers to crunch data without sharing inputs containing private information with one another. Data analysis techniques that combine this form of computation with federated

learning and differential privacy could enable multiple data owners to do encrypted training and prediction, leading to insights that could precipitate breakthroughs in health, education, trade, and other data-driven fields. Promising applications include helping trade agencies recover diverted goods or empowering banks to safely share sensitive ledger data to prevent money laundering.

PETs could even help develop technologies that improve the early detection of cancer and other rare diseases by spotting small changes on medical imagery that even an experienced professional might miss. Training a computer to scan images for the minute changes that suggest cancer requires showing it a vast number of images—some with cancer, some without—so that the machine learns which features are significant. In general, the more data processed, the greater the accuracy.

The challenge, however, is twofold: most people rightly value the privacy of their medical information, and most medical data is collected by a variety of hospitals, clinics, and doctors that don't share their files with one another. Owing to privacy protections and fragmented data storage, the best machine-learning classifiers have, until now, trained on only a tiny fraction of relevant medical records. PETs make it possible to train image classifiers using real medical records, without disclosing any private data. Academic researchers are already piloting applications of this type. Many of the most exciting uses of PETs are still in the early stages of their development, so careful investment and policymaking today could have revolutionary implications for global health and well-being tomorrow.

AUTHORITARIANISM ONLINE

Alas, not everyone is interested in safeguarding privacy in the quest to exploit big data. Not only are China and other repressive countries disregarding privacy and individual rights and freedoms in their own processes, but they may also seek to undermine PETs developed by the United States and its allies. China's promulgation of 5G standards, recent proposals for a new Internet protocol that will give authorities more tools to control access to the Internet, and new facial recognition standards in the International Telecommunication Union are troubling cases in point.

Authoritarianism is not the only threat to realizing PETs' humanitarian potential; another is greed. The value and efficiency of digital networks increase exponentially with size. Private corporations could lock in exclusive access to critical data sets, using data monopolies to extract economic rents and political power. In fact, the risks of monopolies on data may be greater with PETs than they are with traditional data infrastructure. Traditional infrastructure involves making copies of data and spreading them around a marketplace of buyers and sellers. Since every time a party sells or shares data it sends a copy to a second party, it thereby creates another competitor and works against monopolization. However, if data can be used without being copied, the risk of domain-specific data network monopolization and rent seeking is greater.

To understand how a monopolist could use PETs to sell access to data while retaining exclusive control, consider cancer detection again. An enterprising company could develop a federated network for sharing cancer data and, subsequently, cancer-based artificial intelligence (AI) research, products, and services. If this company owned a network with access to the first five percent of the world's data on a particular disease, the next-largest data set would already be at a significant disadvantage when it comes to research and development. When another party wanted to market an additional 10,000 cancer images to researchers, it would naturally go to the biggest network. While PETs remain nascent, one question looms large: Who will own and operate these networks, those who champion the public good or those who seek private gain?

Just as rideshare companies have an incentive to fight pitched battles because a company with 60 percent market share has far more power than one with 40 percent, network effects could make monopolistic control of access to data an even more significant public policy challenge. Companies, governments, or other actors could seek to place networks of data needed to solve public problems into private walled gardens to which access is limited for profit or power. To the extent that hardware products go alongside any applications of PETs, they will create even higher barriers to entry. And because centralized control of PETs can either reinforce or detract from commitments to individual rights and privacy, market share has political consequences, too.

DEVELOPING GLOBAL STANDARDS

PETs are ripe for coordination among democratic allies and partners, offering a way for them to jointly develop standards and practical applications that benefit the public good. At an AI summit last July, U.S. Secretary of State Antony Blinken noted the United States' interest in "increasing access to shared public data sets for AI training and testing, while still preserving privacy," and National Security Adviser Jake Sullivan pointed to PETs as a promising area "to overcome data privacy challenges while still delivering the value of big data." Given China's advantages in scale, the United States and like-minded partners should foster emerging technologies that play to their strengths in medical research and discovery, energy innovation, trade facilitation, and reform around money laundering. Driving innovation and collaboration within and across democracies is important not only because it will help ensure those societies' success but also because there will be a first-mover advantage in the adoption of PETs for governing the world's private data–sharing networks.

Accelerating the development of PETs for the public good will require an international approach. Democratic governments will not be the trendsetters on PETs; instead, policymakers for these governments should focus on nurturing the ecosystems these technologies need to flourish. The role for policymakers is not to decide the fate of specific protocols or techniques but rather to foster a conducive environment for researchers to experiment widely and innovate responsibly.

Democracies should identify shared priorities and promote basic research to mature the technological foundations of PETs. The underlying technologies require greater investment in algorithmic development and hardware to optimize the chips and mitigate the costs of network overhead. To support the computational requirements for PETs, for example, the National Science Foundation could create an interface through CloudBank and provide cloud compute credits to researchers without access to these resources. The United States could also help incubate an international network of research universities collaborating on these technologies.

Second, science-funding agencies in democracies should host competitions to incentivize new PETs protocols and standards—the collaboration between the United States and the United Kingdom announced in early December is a good example. The goal should be to create free, open-source protocols and avoid the fragmentation of the market and the proliferation of proprietary standards. The National Institute of Standards and Technology and other similar bodies should develop standards and measurement tools for PETs; governments and companies should form public-private partnerships to fund open-source protocols over the long term. Open-source protocols are especially important in the early days of PET development, because closed-source PET implementations by profit-seeking actors can be leveraged to build data monopolies. For example, imagine a scenario where all U.S. cancer data could be controlled by a single company because all the hospitals are running their proprietary software. And you have to become a customer to join the network.

Third, democratic policymakers must write legislation that facilitates the responsible adoption of these technologies and encourages the development of appropriate safeguards. Two areas ripe for immediate focus are data consortia agreements and safe harbor laws. PETs require new approaches to data sharing that may involve adjusting existing data protection laws and considering new forms of oversight. To promote innovation, legislatures may need to create regulatory sandboxes, areas where technologists experimenting with applications are temporarily exempt from regulation. With the right regulatory frameworks, adequate liability protections, and meaningful incentives, researchers will develop ever better practical applications.

To test open protocols and demonstrate data security, democratic governments could work together under the auspices of the UN, the G-7, or the Global Partnership on Artificial Intelligence (an initiative hosted by the Organization for Economic Cooperation and Development). Researchers in democracies should also create digital environments to test the analytic utility of PETs. By comparing realistic analytic tasks using this new approach with tasks using traditional data approaches, researchers could prove the value of PETs and overcome the bureaucratic hurdles that new technologies often face.

Fourth, democracies will need to develop new processes for reviewing the ethics of data sharing. If 40 universities across 25 countries are collaborating to analyze data about educational interventions for six-year-olds with autism in 300 jurisdictions, there must be a way to ensure that the collaboration can be expeditiously carried out while protecting

privacy. That means not separate ethics reviews by 40 universities and 300 jurisdictions but some sort of process that is time-efficient and credibly upholds common standards. Just as the Internet prompted the creation of the Internet Corporation for Assigned Names and Numbers, democracies will need to create new intergovernmental bodies above national agencies to govern PETs.

Finally, democracies must have frank and open conversations with their citizens about the appropriate uses of PETs and the stakes for their democratic development. Policymakers should challenge the presumption that there is an irreducible tradeoff between data analysis and privacy—the notion that although more data may improve the accuracy of models for the public good, the cost to privacy is too high. In fact, the status quo is too costly to ignore. Data anonymization techniques offer the simulacrum of privacy, even though enterprising bad actors can easily recombine data records, de-anonymize them, and sell the product as market intelligence to profit-hungry businesses. Although there is no substitute for federal data protection laws, PETs are an important arrow in the quiver for advocates of both privacy and technological innovation whose focus is the public good.

In the years since the Internet first took off, the techno-triumphalism of the 1990s gave way to the sober realization in the 2010s that authoritarian governments can easily use digital technologies for repressive ends. Democracies today have the opportunity to chart a different course with emerging technologies, unburdened by false hopes that they are inherently liberalizing or the false choice that data privacy and data analysis are implacable foes. The operating system of the digital economy needs an upgrade. The question is whether democracies will install the base software for digital societies or whether they will cede that task to others less interested in trust, openness, integrity, and privacy.