Artificial Intelligence/Machine Learning & the Privacy Tech Renaissance

Rachel Zhao (Investor) & Lorenzo Ligato (Intern)
PayPal Ventures

The opinions expressed in this blog are solely the authors' and do not reflect the views of PayPal.

The Essential Role of Privacy: Why It Matters

Digital privacy is crucial for both consumers and enterprises. Consumers want to protect their online presence and personal data by seeking control over who accesses their information and how it is used. For enterprises, safeguarding customer privacy and sensitive data is essential for maintaining trust, ensuring regulatory compliance, and staying competitive in an increasingly data-driven market.

The rise of artificial intelligence (AI) has further amplified the importance of privacy, introducing new challenges and opportunities. AI accelerates the collection, exchange, and synthesis of data at unprecedented speeds, which can increase the risk of data breaches and unauthorized access. As enterprises embrace AI and machine learning (ML) technologies, there is a pressing need for enhanced privacy measures. This shift introduces new opportunities for innovation in privacy-enhancing technologies (PETs), paving the way for a new generation of privacy tech solutions that can help manage and mitigate these evolving risks.

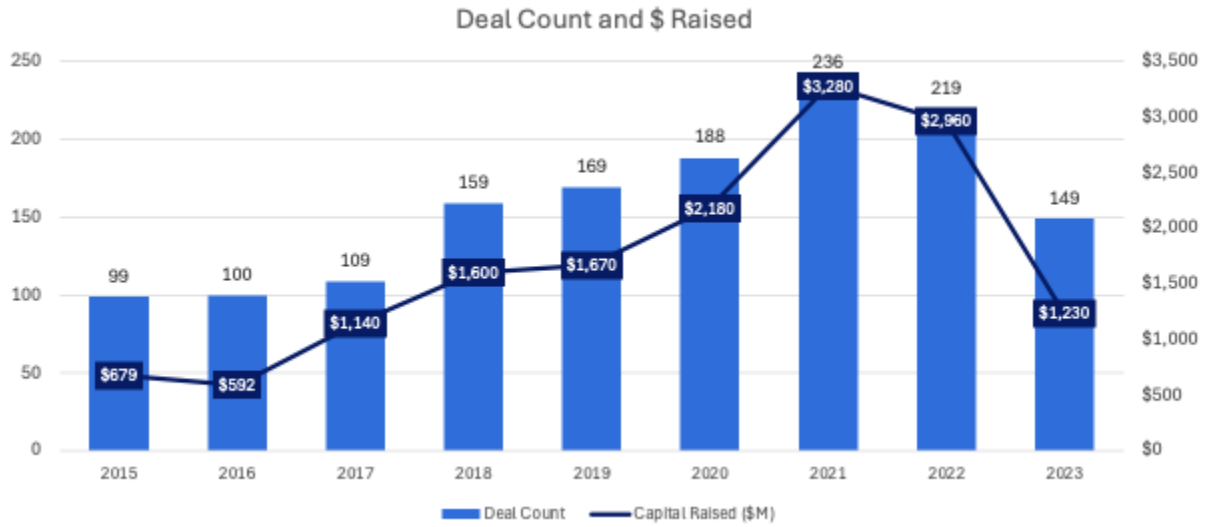
The Past: A Brief Historical Overview of Privacy

Increasing privacy concerns amid technological innovation is not a new development. However, the concept of privacy has undergone significant transformation through the ages. In the digital age, the internet and social media have dramatically reshaped the privacy landscape, spawning unprecedented levels of data collection, surveillance, and information sharing.

The launch of Apple's App Store in 2008 marked a pivotal moment, turning mobile devices into powerful tools for processing data and enabling the growth of apps sharing personal information. Over the next decade, there was a proliferation of startups developing apps focused on both enterprise and consumer privacy. On the enterprise side, companies like OneTrust and BigID emerged to address data protection challenges, while consumer-oriented solutions like Signal, Brave, and DuckDuckGo offered new ways for individuals to safeguard their privacy.

In addition, fundraising activity in the privacy tech space has increased significantly over the past decade and has remained resilient across cycles.

Privacy Tech Fundraising Activities



(Source: Pitchbook)

The Present: Privacy in the Age of AI

The balance between privacy and personalization poses important considerations. Consumers have the choice to share more of their personal data to receive more personalized recommendations, or opt to withhold it and forgo the advantages that analytics can offer in the age of AI. For enterprises, gathering and analyzing customer data can yield insights that bring great advantages for businesses, but they must also stay mindful of the potential risks of data breaches and protect customer data at all costs. After all, the consequences include paying fines and losing trust and reputation.

Below are three privacy-related risks that are specific to AI[1]:

1. *Data Persistence*: Data exists longer than the human subjects who created it.
2. *Data Repurposing*: Data is used beyond its original intended purpose; machine learning algorithms can profile individuals based on otherwise unconnected data points, thus revealing private information.
3. *Data Spillovers*: Data collected on people who are not the target of data collection.

These risks have raised significant privacy concerns regarding the use of AI. Negative perceptions of AI are closely linked to privacy issues. While consumers are excited about the potential of GenAI, 57% of them agree that AI poses a significant threat to their personal privacy.[2] Similarly, businesses remain wary of the privacy risks of machine learning algorithms: 48% of Fortune 1000 technology execs say privacy & data leaks are the top concern when it comes to GenAI. [3]

Privacy Regulatory Guidelines

An additional tailwind in support of privacy tech is the ever-evolving regulatory framework around AI/ML. Unlike the mobile revolution, authorities across the globe are being proactive in regulating AI. Thus, venture capital investors and startups are riding this tailwind to catch more opportunities.

In the United States today, there is no single federal legislation regulating AI. However, the federal government has put out the following guidelines to date:

- October 2022 - Blueprint for an AI Bill of Rights[4]
  - Non-binding, but it outlines five principles – including data privacy – to govern the development and deployment of AI in both the public and private sectors.
  - The blueprint calls for "data minimization" in model training and recognizes the role of individual rights and user control over personal data.
- October 2023 - Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence[5]
  - The Executive Order tasks the Office of Management and Budget with updating guidance for federal agencies to conduct privacy impact

[1] https://www.nber.org/system/files/chapters/c14011/c14011.pdf

[2] https://iapp.org/resources/article/consumer-perspectives-of-privacy-and-ai/

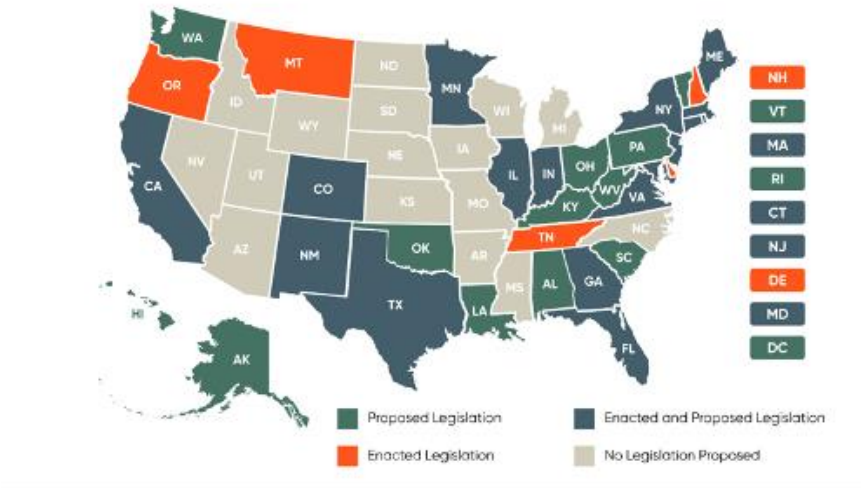
[3] chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.pagerduty.com/assets/whitepaper-generative-ai-survey.pdf

[4] chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf

[5] https://www.nist.gov/artificial-intelligence/executive-order-safe-secure-and-trustworthy-artificial-intelligence

assessments to mitigate risks posed by AI. In addition, the Executive Order directs federal agencies to use privacy-enhancing technologies to protect personal information.

In the absence of federal legislation, multiple U.S. states and municipalities have proposed or enacted laws to regulate AI and AI-related privacy risks, please find the graph below.[6]



In Europe, there are a number of laws regulating artificial intelligence:

- The EU AI Act was published in July 2024, making it the first comprehensive AI regulation by a major regulator. [7]
- Other EU laws and regulations such as GDPR also have implications for AI, such as privacy by design throughout the AI lifecycle, transparent data processing, consent management, etc. [8][9]
-

The Future: Emerging Privacy-enhancing Technologies and Use Cases

Privacy-enhancing Technologies (PETs) are a set of technologies that allow firms to extract, manipulate, and analyze data without compromising individuals' privacy. These technologies can potentially address the dilemma between data privacy and analytics.

[6] https://www.bclplaw.com/en-US/events-insights-news/us-state-by-state-artificial-intelligence-legislation-snapshot.html

[7] https://artificialintelligenceact.eu/

[9] https://securiti.ai/impact-of-the-gdpr-on-artificial-intelligence/

PETs are applied across the model development lifecycle – from ML training to monitoring and observability. Overall, there are 5 main use cases for PETs in AI/ML:

1. **Confidential Computing:** Confidential computing is a cloud computing technology that protects data "in use" (vs. "at rest" or "in transit"). Confidential computing uses a hardware-based architecture – a secure coprocessor inside a CPU called a trusted execution environment (TEE) – with embedded encryption keys. The contents of the TEE are accessible only to authorized programming codes and are invisible to anything or anyone else, including the cloud provider. If a bad actor attempts to access the keys, the TEE denies access to the keys and cancels the computation.[10] [11]

2. **Homomorphic Encryption:** Homomorphic encryption allows different parties to share and analyze encrypted data with one another, without ever knowing what the underlying unencrypted data was. Through homomorphic encryption, the original data (plaintext) is converted into an unintelligible form (ciphertext) that maintains the same structure as the original dataset. Therefore, the encrypted data can be analyzed and worked with as if it were still in its original form. [12] [13]

3. **Federated Learning:** Federated learning is a technique to train AI models on multiple distributed datasets, without the datasets being shared by their owners. Under federated learning, each researcher downloads the model from a datacenter in the cloud, trains it on their private data, and then summarizes and encrypts the model. The model updates are sent back to the cloud, decrypted, averaged, and integrated into the centralized model.[14] [15]

4. **Differential Privacy:** Differential privacy introduces statistical noise – slight alterations – to mask datasets. The noise hides identifiable characteristics of individuals, but it's small enough to not materially impact the accuracy of the model. Before queries are permitted, a privacy "budget" is created, which sets

---

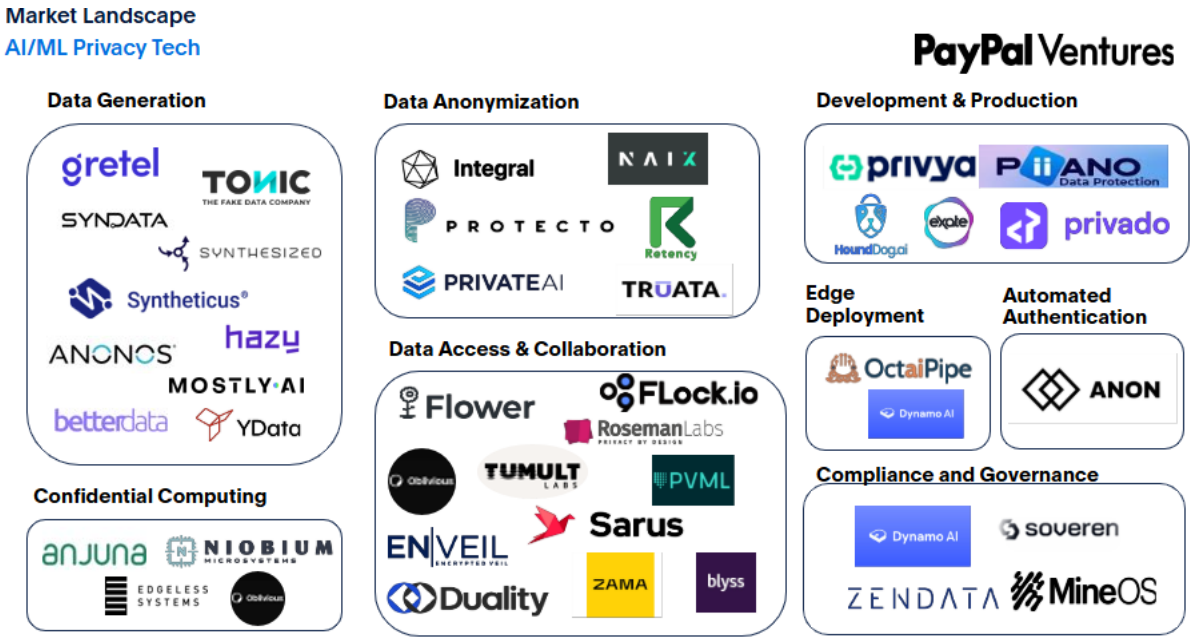
- [10] https://www.fortinet.com/resources/cyberglossary/confidential-computing
- [11] https://www.ibm.com/topics/confidential-computing
- [12] https://www.splunk.com/en_us/blog/learn/homomorphic-encryption.html
- [13] https://www.ibm.com/topics/homomorphic-encryption
- [14] https://blogs.nvidia.com/blog/what-is-federated-learning/
- [15] https://research.ibm.com/blog/what-is-federated-learning

limits on the amount of information that can be extracted from the data. Once that budget has been used up, additional queries are prevented.[16]

5. **Synthetic Data:** Synthetic data generation leverages machine learning methods to produce artificially manufactured datasets from existing, real-world datasets. The synthetic dataset resembles but is not a replica of the underlying real-world data. Therefore, it can be used as a stand-in for testing.[17]

Privacy Market Map

As we enter the AI era, privacy concerns are becoming increasingly critical. Numerous startups are proactively utilizing PETs and other cutting-edge solutions to help consumers and enterprises enhance privacy without compromising the user experience. PayPal Ventures has strategically categorized these startups in the below landscape.



(Sources: Pitchbook, PayPal Ventures)

What Makes us Excited?

With the proliferation of AI, privacy has become even more crucial and must be prioritized.At PayPal Ventures, we are very interested in new, innovative privacy technologies that enhance privacy without sacrificing the customer experience.

- [16] https://privacytools.seas.harvard.edu/files/privacytools/files/pedagogical-document-dp_new.pdf
- [17] https://research.ibm.com/blog/what-is-synthetic-data

With the advancement of AI technology and the rise of Generative AI, data has become increasingly important for training models and generating desired outcomes for businesses. However, this reliance on data also puts privacy at greater risk, as data serves as the foundation for these technologies.

The balance between data privacy and analytics raises the question: can we achieve both without making compromises? We continue to explore startups that offer advanced analytics and insights without sacrificing the protection of sensitive data.

PayPal Ventures Privacy Portfolio Spotlight – MineOS.ai

PayPal Ventures is invested in the consumer privacy space via MineOS.ai. MineOS is recognized as a cost-effective data governance platform for managing privacy, security, and compliance. It started with a consumer product that can help customers track their online footprint, and then organically expanded to serving enterprise customers with privacy automation.

If you are a founder working on exciting privacy technology, please let us know in the comments below!