Secure Data Collaboration Using Privacy-Preserving Technologies

Abhi S Kumar[1], Chinmay C.S[2], Siddesh K.R[3], and Dr. Chitra B T[4]
[1,4] *Department of Industrial Engineering and Management, R V College of Engineering, Bengaluru, India*
[2,3] *Department of Information Science and Engineering, R V College of Engineering, Bengaluru, India*

Abstract*—The proliferation of data-driven business models has led to significant privacy and intellectual property (IP) issues, especially for early-stage startups relying on user data and new sources of insights. Privacy-preserving technologies (PPTs) such as federated learning (FL), differential privacy (DP), homo- morphic encryption (HE), and secure multi-party computation (SMPC), enable technical processes for data utilization while preserving the individual's privacy and IP. The purpose of this paper is to examine the current state of PPTs and provide a conceptual framework for their role in startup settings. The paper draws on the latest literature on the topic and shows how FL and DP allow for collaborative training of models between organizations without the sharing of centralized data [1], [6], how HE enables computations to be performed on encrypted data [4], [13], and the means in which SMPC can build analytics without sharing the private inputs [16], [18]. We then assessed their applications in startup ecosystems and IP management with situations such as collaborative product development and decentralized data marketplaces where FL or HE could be leveraged to extract value from mutual insights without sharing confidential data or trade secrets [8], [20]. Finally, we researched the legal and economic motivations for adopting the technology such as fulfillment of GDPR and CCPA requirements which contributed to adoption by emerging companies [7], [12]. We illustrate how PPTs enable the opportunity for users to build trust and have more regulatory alignment to committing to privacy- centric systems [14], [15]. Practical cases, such as healthcare analytics or monetizing genomic data demonstrate the tradeoffs between privacy, utility, and computational costs [5], [19]. Finally, we note the remaining challenges (e.g., technical scalability, regulatory ambiguity, and economic disincentives), and outline promising future directions (e.g., regulatory standardization, ethical design, and best practices for specific industries) [2], [3], [10]. Overall, we show how PPTs can enable innovation while reinforcing privacy protection for both customers and the firms' intellectual property.

***IndexT e r m s* —Privacy-preserving technologies, federated learning, differential privacy, homomorphic encryption, secure computation, intellectual property rights, entrepreneurship.**

I. INTRODUCTION

The digital economy is built on data, a key resource for in- novation and competitive advantage. Particularly for startups,
user data is often collected or acquired in many situations to train AI models, add value and insights to existing services, and create new products. At the same time, individual people and regulators require privacy protections, and companies want to protect their proprietary information (e.g., trade secrets, algorithms). This presents a dilemma: how can entrepreneurs continue to extract the most value from their data, while minimizing privacy violations and revealing their proprietary IP? PPTs offer to help solve this problem [7].
PPTs (privacy-enhancing technologies) allow machine learn- ing or analytics to be done on sensitive data without displaying that data. Examples of these technologies include federated learning (which enables models to be trained on a variety of devices or organizations without concentrating raw data), differential privacy (which allows information to be hidden by adding noise to sensitivity outputs so that individual data is not reverse-engineered), homomorphic encryption (which enables computations on encrypted data), and secure multi-party computation (which allows the calculation of a function from private inputs) [2].
These tools may enable newer startups to work

together, or make use of local data without disclosing it, and some (such as federated learning) hold promise for collaborating on research or developing products: Berlin-based startup Apheris utilizes FL to enable chemical manufacturers to train federated models on commercially sensitive data, while keeping the data hidden from competitors [20].

This report aims to conduct a survey of existing PPTs, explore their relevance to entrepreneurship and intellectual property rights (IPRs), and propose some frameworks for their use in significant startup applications. We will summarize research related to PPTs, and to privacy in AI. In addition to synthesizing the findings from these studies, we will also look at the legal/economic considerations for the startup sector of various uses of data including issues (e.g., how regulation on data privacy, models to leverage data, etc.) have implications for startup firms. We also will outline hypothetical use cases and methodological issues for implementing PPTs into entrepreneurial endeavors. In this regard, we emphasize both technical possibilities and social/legal implications, so we can provide a holistic view of how ventures can ethically harness PPTs to protect user privacy, and their own IP, while conducting data-driven innovation [12].

II. BACKGROUND AND RELATED WORK

A. Privacy-Preserving Technologies

Modern PPTs can be categorized based on their techni- cal approach. Federated Learning (FL) is an example of a distributed machine learning approach where each participant (e.g., company or device) trains a local model on their private data, and each participant shares only the model updates (gradients) to the data center for aggregation. This keeps private data local and at-risk of privacy leakage low [20].

For example, Ju et al. demonstrate an FL-based 'federated prediction model' for stroke risk that multiple hospitals can train as a group model without sharing patient records. They demonstrated that the 'federated' model improved predictive accuracy by 10%–20% over the models trained by individual hospitals and did so while keeping all sensitive data in-house [1].

The Apheris platform also verifies that federated learning allows for AI training across entities 'without the data ever leaving the control of the data owner,' maintaining both privacy as well as IP [20].

Differential Privacy (DP) is an established statistical method that limits the potential privacy loss when a research team re- leases aggregate information. A differentially private algorithm will act via adding some calibrated random noise to data so that an observer cannot know whether any given individual's data was used to produce the output. This could be applied during learning (for example, adding noise to gradients), or query answering. Wang et al. present FinPrivacy, a system detailing how research teams that include original fingerprint data could apply low-rank approximation and Laplace noise to protect their fingerprint data, while returning a quality match under ϵ-differential privacy [7].

This example represents one of the trade-offs involved in

Differential Privacy, maintaining privacy guarantees at the expense of utility in some contexts. For example, in en- trepreneurial contexts, some potential utilities of DP may allow startups to publish limited analytics/pivot tables or market their service while still maintaining compliance to privacy regulations, e.g., GDPR, and without revealing the underlying data of individuals [6].

Homomorphic Encryption (HE) refers to the ability to perform

arbitrary functions/computations on encrypted data. In this approach, data owners encrypt their original data and send ciphertext data to a server to perform some computations, i.e., the server receives encrypted data as inputs to a neural network inference. The server returns a corresponding ciphertext result to the client (the same output that would have been returned if plaintext data had been used), while guaranteeing no under- lying plain text data was revealed [4].

The overarching premise aims to safeguard user data, but conversely, the model IP may also be preserved due to the unpublished model when it was observed during diffusion. Jin et al. suggest FedML-HE, an FL architecture that enables the use of HE to do a secure federated model update aggregation; with these methods significantly improving computation per-

formance (up to a 40× reduction in overhead when applied only to the sensitive parameters) [4].

Secure Multi-Party Computation (SMPC) is a general crypto- graphic method to allow multiple entities to jointly compute a function without revealing their

inputs. For example, two companies could jointly analyze total customer behaviour without having to disclose to each other who their respective customers are. SMPC gives the opportunity to compute the joint function based on the total data collaboratively while providing complete privacy of the data [16].

There are other PPTs which are being investigated, including trusted execution environments, zero-knowledge proofs (ZKP), and private set intersection. The aforementioned example of combining FL and ZKP are examples of topics that our community is actively researching. The work done by Jin et al. [5], which allows confirmation for updates to clients without having access to the update data, is another example of merging FL and ZKP.

B. Privacy in Machine Learning and IoT

A growing body of research has identified scopes of overlap between PPTs and other forms of ML. Rashid and Yasin have organized an extensive taxonomy of privacy-preserving prac- tices in deep learning through reviewing, among others, FL, DP, HE, collaborative protocols, and the modality of sensitive data one might have (such as medical or textual/globally), which work and can be used [18].

Regarding IoT health applications, the paper by Vijayakumar et al. presents the notion of a privacy-preserving federated learning (FL) system in which patients' data are collected from edge devices and aggregated using homomorphic encryption so that no raw medical data is transmitted from the local devices [17].

Work in finance and biometrics has similar results, empha- sizing that using DP, HE, or SMPC means data on private

financial or biometric data can be used for model training and prediction [8].

C. Economic and Legal Context

The technological adoption of privacy-preserving techniques (PPTs) is also influenced by economics and legal factors. Data protection legislation, such as the EU's General Data Protec- tion Regulation (GDPR) and California's Consumer Privacy Rights Act (CPRA), has prompted startups and enterprises to scrutinize how they manage their data. Martin et al. [7] found that GDPR restricted startup innovation and incentivised innovation. For some firms, compliance experience was a hindrance as compliance costs were an impediment, while other firms leveraged privacy as their competitive advantage. Kantarcioglu et al. also develop a game-theoretic model on when firms are willing to invest in privacy-preserving tech- nology (PPTs), concluding that high valuation of privacy by customers and substantial regulatory sanctions make investing in PPTs beneficial [12].

While personal data is not conventionally recognized as in-

tellectual property, startups often view user data and model parameters as proprietary and assert ownership with con- tracts and digital rights [10]. Starting in regards to protecting consumer privacy and business-valued IP assets, encryption, access control, and anonymization will be fundamental [10]. Secondly, since data is commoditized, startups are distilling and monetizing data-driven services with privacy-compliant methods like encrypted analytics or even federated analytics [14].

III. METHODOLOGY

A. Privacy-Preserving Data Collaboration Model

We proposed a conceptual model for startup contexts, to work together on data analytics or a machine learning (ML) model collectively, maintaining user privacy, along with proprietary intellectual property (IP). The framework allows for collaborative capabilities while not sharing raw data, or exposed proprietary algorithms.

1) *Data Governance and Preprocessing*: Collectively, ev- eryone in the collaboration will likely categorize data by sensitivity and use cases, e.g., user records, business-value attributes, proprietary features of an ML model. Preprocessing of data includes processes for managing consent of users, where applicable, anonymization of data where needed, and the applying of privacy methods, like local differential privacy or even encryption [6].

2) *Local Enclaves*: Data is still primarily isolated in local, safe environments: on-device or institutional infrastructure. On the structured data (e.g., categorical or numerical data) side, the data is encrypted, or a local DP (Differential Privacy) mechanism is used. On the unstructured data (e.g., images or textual data) side, one's ability to create synthetic data, or employ obfuscation techniques, will provide protection under expectations of privacy

[13].

3) *Collaborative Protocols*: The framework employs several collaborative protocols. First, Federated Learning involves each participant initializing a common model architecture and completing local learning on decentralized data at their location. Rather than sending the data itself, they share only updates (as noisy updates) with the central aggregator [1], [4]. Second, homomorphic aggregation allows the aggregator to combine the updates with Homomorphic Encryption (HE). Since all updates are encrypted end-to-end, no participant, including the aggregator, will be able to determine the original inputs from the gradients shared across [4]. Third, differential privacy is achieved by participants or the aggregator creating differential privacy from the model updates by adding cali- brated noise to the gradients or the weights. This provides a quantitative measure of privacy (ϵ-differentially privacy) to ensure that no single record or observation informs the model output [6]. Fourth, iterated rounds involve repeating the federated averaging process for several rounds of com- munication. As noted by Chen et al., this process may attain model performance approximating the model developed from centralized training (i.e., 97.3% vs. 98.2% accuracy) while maintaining data locality and in accordance with regulatory requirements [6].

4) *Secure Computation*: The structure supports ad-hoc privacy-

preserving analytics in accordance with several protocols and services. First, Secure Multi-party Computation (SMPC) pro- tocols allow for joint analytics or predictions to be accom- plished over distributed data, while ensuring each participant maintains control of their private input [16]. Second, Trusted Computing Environments (TEE) provide a secure and isolated space in hardware for processing encrypted inputs, which is useful for high throughput inference where software alone is not effective [15]. Third, Zero-Knowledge Proofs (ZKP) allow entities to prove that model updates and steps taken in computation were correct without revealing the underlying data itself. Jin et al. successfully incorporated ZKP into federated learning for additional verifiability in adversarial spaces [5].

5) *IP Controls*: For startups that use privacy-preserving tech-

nologies (PPTs) to help secure their proprietary algorithms, trained models, and sensitive datasets, intellectual property (IP) protection is essential. The recommended model involves both legal and technical means of protecting IP rights while creating an environment for collaborative data analytics. First, technical means to protect IP involve storing the final trained model, which has IP value, in secure locations, i.e., in Trusted Execution Environments (TEEs) or private cloud environ- ments. The startup uses homomorphic encryption (HE) on the 'weights' defined in the model during inference for prediction so that no one else knows the weights or the architecture of the model. This will prevent anyone from accessing the trained model, by preventing representation of the model from being visible and preventing reverse engineering as they communicate through encrypted standard queries available to the model [4], [13]. For example, a startup selling Machine Learning as a Service (MLaaS) can accept user queries that are encrypted to obscure its model parameters in accordance with the principles of trade secret protection as discussed in [13]. Second, legal IP frameworks require collaborators to execute comprehensive legal agreements that clearly outline owner- ship, rights of use, and licensing arrangements for models or insights created within a collaboration. The legal framework is based on intellectual property or IP law, including patent law (e.g., any new algorithms), copyright law (e.g., any software application of the algorithms), and trade secret law (e.g., the model architecture and proprietary data sets) [10]. In the US, see the U.S. Patent Act (35 U.S.C.), and in the EU reference European Union's Directive on Trade Secrets (2016/943). The legal agreements facilitate clear clarified ownership of IP in cooperative or collaborative environments. Further, emerging startups may have NDAs (non-disclosure agreements) in place to protect any proprietary information shared in the context of treating at arms-length in terms of federated learning (FL) or secure multi-party computations (SMPC) [10]. Third, IP rights in data note that personal data does not typically conform to a legal definition of IP. Nonetheless, startups in this space do orient curated datasets (typically based on personal data) as proprietary assets. The framework protects IP rights associated with data

through monitoring access and use permissions and then through encryption and anonymization during FL, while complying with the provisions of the World Intellectual Property Organization (WIPO), who recognize some forms of data as a legitimate form of IP where the ownership and rights associated with intra-data any kind of contractual protection [20]. Moreover, new legal discussion around quasi- IP rights for personal data from the World Intellectual Property Organization (WIPO) may permit individuals to license or monetize their data and would further establish PPTs in secure data marketplaces [10].

6) *Data Monetization*: If monetization was an organizational goal, the framework permits utilization for decentralized data marketplaces. First, smart contract basis allows data providers and consumers to interact using contracts on a blockchain (similarly to the agreements in the previous section). Any transactional agreement (e.g., for a service for analytics) can be supported and are all auditable and verifiable while actual data remains off-chain and private [16]. Second, incentivized privacy sharing mirrors how Nebula genomics enables the user to retain ownership of the user's data while permitting en- crypted access for the computation. Incentives can be provided to providers (e.g., tokens, revenue share) without knowledge of the identity of the user [19].

B. Experimental Design

To demonstrate the feasibility of the framework, the ex- periment should be structured as follows: 1) *Data Partition*: Suppose that there is a consortium of n startups that each hold a different part of user data. The centralized baseline model is trained on the integrated dataset. Other startups ran FL like vanilla FL or FL that uses a specific attribute of privacy-preserving technologies (PPTs).

2) *Privacy Budget*: Each participant who applies differential privacy (DP) method specifies a privacy budget of $\epsilon = 1.0$—very strong protection setting. Noise adds to the model gradients (updates) during FL aggregation [6].

3) *Summary of Performance Evaluation*: There are several things to compare. This includes model accuracy, training time, communication cost, and convergence time. These data are collected for three configurations—Intellectual Property (IP) Protection: The objective is to understand the informa- tion leakage of the model parameters across the different approaches. In the HE FL with DP approach, the aggregator will not ever see the decrypted gradients, so there is zero exposure to proprietary updates [4].

4) *Economic Model*: Ultimately, a cost-benefit analysis seeks to estimate the monetary impact of employing privacy- preserving technologies (PPTs). Generally, the costs are in- frastructure implementation (e.g., homomorphic encryption (HE) computations), service pricing premiums for protecting privacy, and earnings from data-sharing business models or federated monetization. Available preliminary estimates from previous work suggest the possible integration of differential privacy (DP) may diminish model performance by a small amount ($\approx$ 0.9%) when compared to traditional centralized learning, as indicated in Chen et al. [6]. While HE costs are significant due to increased computation and communication load, they ensure compliance with regulations and protect user and business-specific privacy [4]. As demonstrated in Nebula's case study, encrypted query models may allow for monetization without degrading data ownership, or create business models to monetize data without actual data ownership [19].

IV. RESULTS AND DISCUSSION

From our conceptual framework, startups can achieve good utility-like outcomes while preserving privacy and intellectual property (IP).

Tradeoffs Between Utility & Privacy: Similar to previous work, there will be a performance tradeoff as we first implement privacy-preserving technologies (PPTs) in federated learning (FL) applications. For instance, Chen et al., studying federated learning (FL) systems using differential privacy (DP) and homomorphic encryption (HE) for data privacy, achieved a 97.3% accuracy federated learning (FL) system compared to a centralized training FL system accuracy of 98.2% on the MNIST dataset, there was a very small performance reduction ($\approx$ 0.9%) opted into a much larger performance/privacy advan- tage [6]. Additionally, Ju et al. recognized that collaborative stroke prediction models that trained across hospitals using FL increased performance by

10–20% compared to a model trained with data of isolated hospitals in several hospital con- texts [1]. Our different mixture hybrid learning FL structure should still establish performance that would be comparable to the centralized models with understood overhead costs for computation and noise added for differential privacy. Jin et al. also demonstrated that limited use of HE within their FedML- HE system decreased computing overhead when encrypting
and established reduced computing overhead annually between 10× and 40× depending on model architecture, which made HE-augmented FL profound and relatable [4].

Privacy Guarantees: Incorporating formal privacy guarantees represents a strong layer of data protection. One recommen- dation when using $\epsilon = 1.0$ for DP and adding Gaussian noise to model updates is to lower the probability of successfully inferring membership or re-identifying records [6]. For exam- ple, although an adversary could again gain access to model updates, the added noise would obscure individual contribu- tions and improve secrecy if model changes are intercepted. Because HE encryption is employed, the aggregator or other users also cannot see the raw gradients or data, as all updates are encrypted end-to-end during the aggregation process [6]. Overall, these two forms of protection can comply with data protection requirements while sustaining an analytical product.

Communication and Computation Costs: PPTs incur compu- tational and bandwidth overheads. Homomorphic operations are slower than similar plaintext operations, and bandwidth costs are incurred by sending encrypted updates. Jin et al. have benchmarked and achieved $\sim 10^{th}$ speedups on ResNet and about 40^{th} speedups on BERT with optimization for HE in FL—but the absolute cost of decentralized HE FL is still higher than plaintext FL [4]. New startups who deploy such systems should expect to pay for extra compute, for instance when using cloud GPU instances, and allocate longer time for training. Good engineering design choices, such as only encrypting parts of the model which contain sensitive information, can help minimize some of these costs while still meeting security requirements.

Model and Data Ownership: In our model, every startup will maintain ownership of the raw data, acting on the principle of 'data never leaves,' which is core to FL design [20]. The global model will be co-owned or protected through legal agreements. Distribution using, HE and SMPC ensures that neither data nor proprietary features will be revealed in the training processes, directly enabling IP protection. Wang et al. also explored how they protect IP by using model encryption or running the model in a secure enclave, while still offering the possibility to provide inference services to clients [13]. For instance, a startup could host models in a secure server, and the input would be encrypted through HE, whilst allowing third parties to request queries without revealing model parameters from the startup or inputs from clients [13].

Legal Compliance: The framework is consistent with pre-
existing data protection regulations (GDPR and CCPA), as it limits exposure to personal data. Martin et al. discovered that privacy laws can serve as barriers and accelerators of innovation, forcing people to rethink architectures, while also enabling unique business models based on privacy compli- ance [7]. In this sense, startups that adopt PPTs can convert regulatory limitations into a marketing advantage. As evalu- ated legally and ethically, privacy-oriented design can build consumer trust and, hence, potentially increase data sharing, customer trust, and, another key, competitive advantage [10]. A 2025 IMF advisory report has identified that privacy technologies improve regulatory compliance, stimulate new market dynamics, and promote territorial or cross-border data collaboration, related to the Pew Research Center Commission on Privacy in the Digital Age, a report and distillation of perspectives about conditions of cooperation [15].

For example, the Nebula genomics platform allows individ-
ual consumers to share their encrypted genomic data for financial compensation while also providing researchers with anonymized findings—this scenario strikes a balance [19].

In conclusion, we believe that enabling PPTs allows for secure and collaborative analytics with a clear promise of privacy. By using federated learning, differential privacy, and homo- morphic encryption, entrepreneurial ventures can collaborate in creating effective models, or can collaboratively deliver better, data-driven services, while upholding user privacy and without disclosing business secrets. The

societal upside is equally substantial—strong privacy protective measures allow a wider number of participants into data ecosystems and hence, greater innovation. As Lucie Arntz from Apheris said, federated approaches can provide 'magical' value by linking datasets that have almost no value independently—something that we have converted into a useful framework [20].

V. APPLICATIONS

A. Collaborative Innovation

Startups often need to collaborate across institutions or share pooled data before they can reach sufficient scale for innovating. Examples include a consortium of health-tech startups or research universities who may develop a disease prediction model together. Under the scheme we propose, the parties could train on site, using their patient records data, and only share encrypted updates using FL. The privacy of individ- ual actors would be honored while improving generalization performance from aggregate datasets that wouldn't have been possible otherwise (e.g., 'different cancers') [1].

Similarly, in other domains such as manufacturing or automo- tive, companies can share their datasets, such as LiDAR scans or vehicle telemetry, and support collaborations where the pace of research of autonomous systems such as self-driving vehicles engines, etc., is accelerated. Apheris has shown how FL can allow companies to collectively and jointly train based on corporate industry data without sharing manufacturing data, which can offer protection to commercially sensitive data [20]; or software startups can use secure multiparty computation (SMPC) to share analysis to identify insights into market trends across organizations without breaching client identity or sensitive competitive data [16].

B. Protecting Intellectual Property

Privacy-preserving technologies (PPTs) span two foci: the protection of user privacy and the protection of startup IP. Combining PPTs and IP laws allows startups to legally comply with the use of privacy technology, while maintaining compet- itive advantages.

1) *Compliance with IP Laws*: Depending on the PPT, tech- nologies such as FL, HE, and SMPC protect startup IP by not exposing proprietary algorithms and datasets. For instance, based on the U.S. Defend Trade Secrets Act (DTSA) of 2016, in the instance where a startup uses proprietary model parameters, it may utilize HE to encrypt updates using FL, allowing trade secrets to remain confidential while engaged in joint training [4]. Additionally, the framework of the EU's Patent Cooperation Treaty (PCT) supports patenting innovative methods to deploy PPTs, such as HE algorithms, to further protect and develop unique intellectual property for startups [20].

2) *Contractual IP Protections*: Legal agreements are impor- tant to define IP rights in collaborations. Agreements will not only define rights to foreground IP (new models or new understandings) but also make provision for background IP (previously owned intellectual property). For example, a health-tech startup using FL to train a model for predicting disease could have agreements that allow them to retain control of proprietary aspects of their model (e.g., how they construct a unique feature-set for training) in situations where they share model updates (e.g., TEE updates are aggregative) [1], [10]. The agreements that are made would most likely comply with the principles of WIPO on IP management in collaborative research [20].

3) *IP Monetization*: Point-to-Point Trust (PPT) contracts allow

startups to monetize their IP assets in ways that do not compromise privacy or ownership. For example, a startup could license its model to make inferences in a TEE, while having users use fully homomorphic encryption (FHE) to generate encrypted queries, and funnel revenues back to them without loss of control over ownership of the model. A good example of this is Nebula Genomics, which allows individual subscribers to share their encrypted genomic data to receive compensation. Nebula also utilizes blockchain and smart contracts to guarantee that IP issues are maintained [19]. Extensions of the model could occur in almost any setting where the startup can build revenue on the encrypted analytics, whilst at the same time reserving ownership of the IP.

C. Data Monetization and Markets

PPTs are essential to new business models that center on data monetization, all while ensuring privacy protections. One form of innovation that has emerged is decentralized data mar- ketplaces. The Nebula

Genomics platform is a more tangible example, allowing individuals to share encrypted genomic data for cash, while organizations run privacy-preserving queries against that data using permissioned blockchain-backed access controls. Thus, neither party is exposing their data or principal in their tradable product [19].

This ecosystem represents a departure from secure data owner- ship norms. Individuals have control over the sensitive health data they share, and researchers effectively receive value from the data through secure computation rather than retaining or having access to the actual data. Startups in finance, marketing, and e-commerce may replicate this alternative access model, allowing them to offer analytics or personalized service based on an encrypted, anonymized contribution. The startups can even monetize aggregated predictions or trending insights, without violating privacy protections or trade secrets.

D. Legal/Economic Incentives

As startups navigate tighter regulatory environments for data, they face obligations after compliance, such as the EU's General Data Protection Regulation (GDPR) and California's Consumer Privacy Act (CCPA), around user consent, access to the data, data portability, and erasure [10]. Startups will be able to credibly demonstrate compliance, including verifiable assurances that individual data are not inferable from aggregate outputs, using techniques like DP and HE [6].

Startups also tend to view their curated datasets and models (even without an IP claim) as being part of a core intellectual property. The use of encryption, SMPC, and federated archi- tecture allows the intellectual property to remain in the sole control of the owner, and at the same time provide collabo- rative analysis or service delivery [10]. WIPO's commentary cited earlier stated that confidentiality and privacy protection can be nested (there is no conflict between privacy protection and IP preservation), and each would reinforce the other in cases with contractual clarity and technology to enforce [20]. PPTs are adopted for varying reasons, including legal compli- ance and economic incentives associated with protection of IP. Data protection laws, including the EU's General Data Protec- tion Regulation (GDPR) and California's Consumer Privacy Act (CCPA), impose requirements and restrictions on how data is handled, and some PPTs fulfill the GDPR and CCPA requirement of ensuring data security through encryption and anonymization [10]. At the same time, IP laws encourage the adoption of PPTs, providing legal protection for proprietary innovations.

1) *IP Law Compliance*: Compliance with GDPR's Article 32

requires organizations to establish protections like encryption, which the HE and SMPC do, in turn allowing it to have some level of assurance it is complying with the GDPR while at the same time protecting its proprietary data [10]. The DTSA and EU Trade Secrets Directive will give potential sources of legal recourse for startups if the proprietary data and/or models used by the startups are appropriated; therefore, the need to establish PPTs will help avoid startup appropriation during collaborative analytics [20]. Startups can utilise these laws to create trust with both consumers and regulators, and the startups can capitalize on this perceived value to establish privacy as a competitive advantage.

2) *Monetary Considerations for Intellectual Property Protec- tion*: Kantarcioglu et al. present convincing evidence showing that where there is a high consumer value of privacy (and heightened regulatory penalties associated with breaches), organizations have strong incentive to make investments in privacy-preserving technologies (PPTs) that rely on intellectual property [12]. When companies are able to protect intellectual property with their PPT, they are more able to entice venture capital investment, as privacy-compliance improves company valuation under frameworks like the environmental, social, and governance (ESG) guidelines. Take, for example, specific startups attempting to use federated learning to protect their proprietary datasets—they can market privacy-compliance to privacy-sensitive investors by showcasing how use of their intellectual property is lawful and documented [12]. Notewor- thy also is the suggestion in the World Intellectual Property Organization (WIPO) 2023 report that innovations arising from use of IP-protected data in PPTs are also used by others as a basis for selling licensed data, which generates new revenue opportunities for companies [20].

VI. CHALLENGES AND FUTURE DIRECTIONS

Privacy-preserving technologies (PPTs) represent significant promise but are hampered by a multitude of technical, le- gal, and social challenges that need to be overcome before PPTs will be used widely, incorporated within emerging entrepreneurial ventures.

Scalability and Efficiency: Cryptographic operations, and in particular fully homomorphic encryption (FHE) operations, are still computationally expensive. The operational costs of systems of federated learning are impressive as even when delivered via optimized frameworks, e.g., FedML-HE which encrypts parameters only on sensitive parameters, are expen- sive in terms of computing power and memory bandwidth (overhead usage that can be considerable particularly with large models like BERT [4]). Furthermore, most early-stage startups do not have the hardware and supporting infrastructure (dedicated GPUs or FPGAs) to meet such processing workload requirements. In this specification focus, future research needs to consider algorithmic compression, hardware acceleration, and the hybridization of encryption. Finally, federated systems require support for diverse federated participants and network conditions, so asynchronous protocols and fault-tolerant ag- gregation need robust support.

Standardization and Interoperability: As described in the

Apheris case, collaboration of shared privacy is limited by a lack of data schema consistency and standardization of the interface. Industry experts say we need protocols for data formatting, ontologies, and secure APIs to facilitate privacy- preserving data sharing across silos [20]. In connection to health care, autonomous mobility, and other sectors of industry that need cross-organization collaboration with their data, regulatory bodies or technical consortiums may need to define the interoperability standards.

Legal and Regulatory Uncertainty: The interface between PPTs and IP laws will require further examination due to complexities caused by the ownership question surrounding use of personal data that is only sometimes considered IP [10]. For example, there is ambiguity as to whether consent for per- sonal data of IP-protected data is fully compliant with GDPR consent requirements under its 'broad consent' provisions even when use is repressive and therefore consumer-consented [20]. Future legal frameworks should specify guidance as to how PPTs operate within IP laws, like the Copyright Act of the United States and the Database Directive of the European Union, for example, and how ownership of personal data rights for datasets that are aggregated or products delivered in the form of trained models could emerge from further privatisation efforts in academic and research settings. The debates at WIPO about quasi-IP rights of personal data could strengthen the relationship between PPTs and IP governance, allowing startups to copyright their data under license [10]. *Economic Incentives*: The creation and usage of PPTs require a considerable investment in terms of legal, cryptographic, and computational capabilities, which may be a major barrier for startups. Additionally, users tend to expect privacy, but generally are unwilling to pay more for it, leading to a 'privacy gap.' Recognizing this misalignment creates possible economic alignment, potentially through privacy certifications, tax credits, or marketplace compensation mechanisms (e.g., to- kens for privacy-respecting data usage). As demand increases for compliant data solutions, the number of venture capital investments in privacy-first startups may therefore increase, especially for enterprises who have a need for secure B2B data pipelines [12], [19].

Security and Robustness: PPTs preserve privacy but have

the potential to introduce new vulnerabilities regarding the integrity of a system. One example is federated learning (FL), which is vulnerable to model poisoning where an adversar- ial client uploads malicious updates. Zero-knowledge proofs (ZKP) are proposed solutions to verify that computations done on the client-side are correct without revealing the data [5]. These are still developing as cryptographic con- cepts, particularly regarding their scalability. However, zero- knowledge proof-based mechanisms are still evolving. Even though differential privacy is inscribed with formal guarantees, its randomized mechanisms could put some accuracy at risk, while exposing majority group bias. Continuing research will be needed to explore these trade-offs and train systems to fight emerging threats such as quantum

cryptanalysis, model inversion, etc. *Ethical and Societal Issues*: Furthermore, technical privacy is not equal to user trust, transparency, or fairness. Methuku et al. argue that ethical frameworks should accompany technical solutions to trigger the responsible use of data in research [14]; for example, some studies using DP may present unequal error rates across demographic groups. Furthermore, if the rationale behind encrypted systems, and thus the protocols themselves, are not understood, users will also distrust the systems. There- fore, public awareness campaigns, as well as user interfaces which communicate the purpose behind processes, are needed to help reduce the cognitive dissonance experienced by users and their understanding of the mathematical constructs that underpin cryptography.

Future Directions: Future work should consider aligning PPTs

with IP legislation to create vanilla arrangements for owner- ship of data and models (for example, by combining zero- knowledge proofs (ZKPs) with FL to allow verifiably feature- less IP protection to startups to provide evidence of their model without requiring the disclosure ownership of parameters proprietary to the model [5]). Furthermore, blockchain-based IP registries could allow for the representation of ownership of models specified in PPTs, demonstrating auditability and compliance with international IP structures like the PCT [20]. Governments and organizations [15], including the Interna- tional Monetary Fund (IMF), are examining how to enact PPTs in service of digital identity and new approaches to IP management, indicating that the PPT role in international IP regimes will likely expand.

VII. CONCLUSION

Privacy-preserving technologies (PPTs) are rapidly chang- ing the way that data-driven companies are doing business. PPTs enable joint learning from data in a distributed en- vironment (i.e., data-driven innovation and analysis, without storing sensitive data in a centralized way leading to greater risks of data exposure), thus allowing a startup to create new value without re-victimizing privacy, while also safe- guarding their intellectual property (IP). We discussed relevant PPTs, namely federated learning (FL), differential privacy (DP), homomorphic encryption (HE), and secure multiparty computation (SMPC) as technologies and their application to entrepreneurship and intellectual property rights (IPR), along the way providing significant technical detail, practical examples, and legal-economic relevance.

The case studies indicated that while PPTs may create some

level of computational and basic deployment complexity, they reveal substantial benefits: improved performance of models obtained collectively [1], compliance with growing global data protection laws (such as GDPR/ARA and CCPA) [10], and finally, privacy and proprietary knowledge protection of data and models [13], [20].

Moving forward, there is a need for continual research to make these technologies operationally effective in startup ecosystems by improving either their cost-effectiveness, ef- ficiency of data processing for these systems, or at minimum, developing better standards for using PPTs and making them less burdensome for purchase. Within a growing ecosystem of regulatory mandates of use and emergence of consumer protectors of privacy (expectations), there is an opportunity to make, or indeed to expedite a movement towards, a version three (in terms of privacy) of ‘privacy by design’. While digital innovation will remain reliant on data as a key element of innovation, PPTs will enable the ability for privacy rights and IP security to develop in alignment with entrepreneurial intent.

REFERENCES

[1] C. Ju, R. Zhao, J. Sun, et al., “Privacy-Preserving Technology to Help Millions of People: Federated Prediction Model for Stroke Prevention,” arXiv preprint arXiv:2006.10517, 2020.

[2] X. Qi and M. Zong, “An Overview of Privacy Preserving Data Mining,” Procedia Environmental Sciences, vol. 12, pp. 134–139, 2012.

[3] I. Habernal, F. Mireshghallah, P. Thaine, et al., “Privacy-Preserving Natural Language Processing,” in Proc. 17th Conf. of the European Chapter of the Association for Computational Linguistics (EACL), Dubrovnik, 2023.

[4] W. Jin, Y. Yao, S. Han, et al., “FedML-HE: An

Efficient Homomorphic- Encryption-Based Privacy-Preserving Federated Learning System," arXiv preprint arXiv:2303.10837, 2024.

[5] Y. Jin, et al., "Zero-Knowledge Federated Learning: A New Trustworthy and Privacy-Preserving Paradigm," arXiv preprint arXiv:2303.18076, 2025.

[6] Y. Chen, Y. Yang, Y. Liang, et al., "Federated Learning with Privacy Preservation in Large-Scale Distributed Systems Using Differential Privacy and Homomorphic Encryption," Informatica, vol. 49, no. 1, pp. 101–112, 2025.

[7] T. Wang, Z. Zheng, A. K. Bashir, et al., "FinPrivacy: A Privacy- Preserving Mechanism for Fingerprint Identification," J. ACM, vol. 37, no. 4, pp. 871–890, 2018.

[8] L. Laishram, M. Shaheryar, J. T. Lee, and S. K. Jung, "Toward a Privacy-Preserving Face Recognition System: A Survey of Leakages and Solutions," ACM Comput. Surv., vol. 57, no. 2, pp. 1–35, 2025.

[9] D. Dhinakaran, S. M. U. Sankar, D. Selvaraj, et al., "Privacy-Preserving Data in IoT-Based Cloud Systems: A Comprehensive Survey with AI Integration," 2024.

[10] World Intellectual Property Organization (WIPO), "WIPO Technology Trends 2023: Artificial Intelligence and Intellectual Property," 2023.

[11] A. S. George and A. H. George, "Data Sharing Made Easy by Technol- ogy Trends: New Data Sharing and Privacy Preserving Technologies," Partners Universal Int. Res. J., vol. 2, no. 1, 2022.

[12] M. Kantarcioglu, et al., "When Do Firms Invest in Privacy-Preserving Technologies?" in Decision and Game Theory for Security (GameSec), Springer, pp. 220–239, 2010.

[13] T. Wang, Z. Zheng, A. K. Bashir, et al., "Privacy-Preserving Mechanisms for Data-Preserving Inference and IP Protection Using Homomorphic Encryption," J. ACM, vol. 37, no. 4, pp. 871–890, 2018.

[14] V. Methuku, et al., "Bridging the Ethical Gap: Privacy-Preserving Artificial Intelligence in the Age of Pervasive Data," Int. J. Scientific Advances, vol. 2, no. 4, pp. 123–133, 2021.

[15] D. C. G. Valadares, et al., "Privacy-Preserving Tools and Technologies: Government Adoption and Challenges," Sensors, vol. 23, no. 3, pp. 2245–2261, 2023.

[16] J. B. Bernabe, et al., "Privacy-Preserving Solutions for Blockchain: Review and Challenges," IEEE Access, vol. 7, pp. 164363–164388, 2019.

[17] P. Vijayakumar, P. Sharma, U. Ghosh, et al., "Homomorphic Encryption- Based Privacy-Preserving Federated Learning in IoT-Enabled Healthcare System," IEEE Trans. Netw. Sci. Eng., vol. 9, no. 1, pp. 22–35, 2022.

[18] N. S. Rashid and H. M. Yasin, "Privacy-Preserving Machine Learning: A Review of Federated Learning Techniques and Applications," Int. J. Sci., vol. 13, no. 2, pp. 89–98, 2025.

[19] D. Grishin, et al., "Accelerating Genomic Data Generation and Facili- tating Genomic Data Access Using Decentralization, Privacy-Preserving Technologies and Equitable Compensation," Blockchain in Healthcare Today, vol. 1, no. 3, 2018.

[20] World Intellectual Property Organization (WIPO), "Intellectual Property and Privacy-Preserving Technologies," 2024.