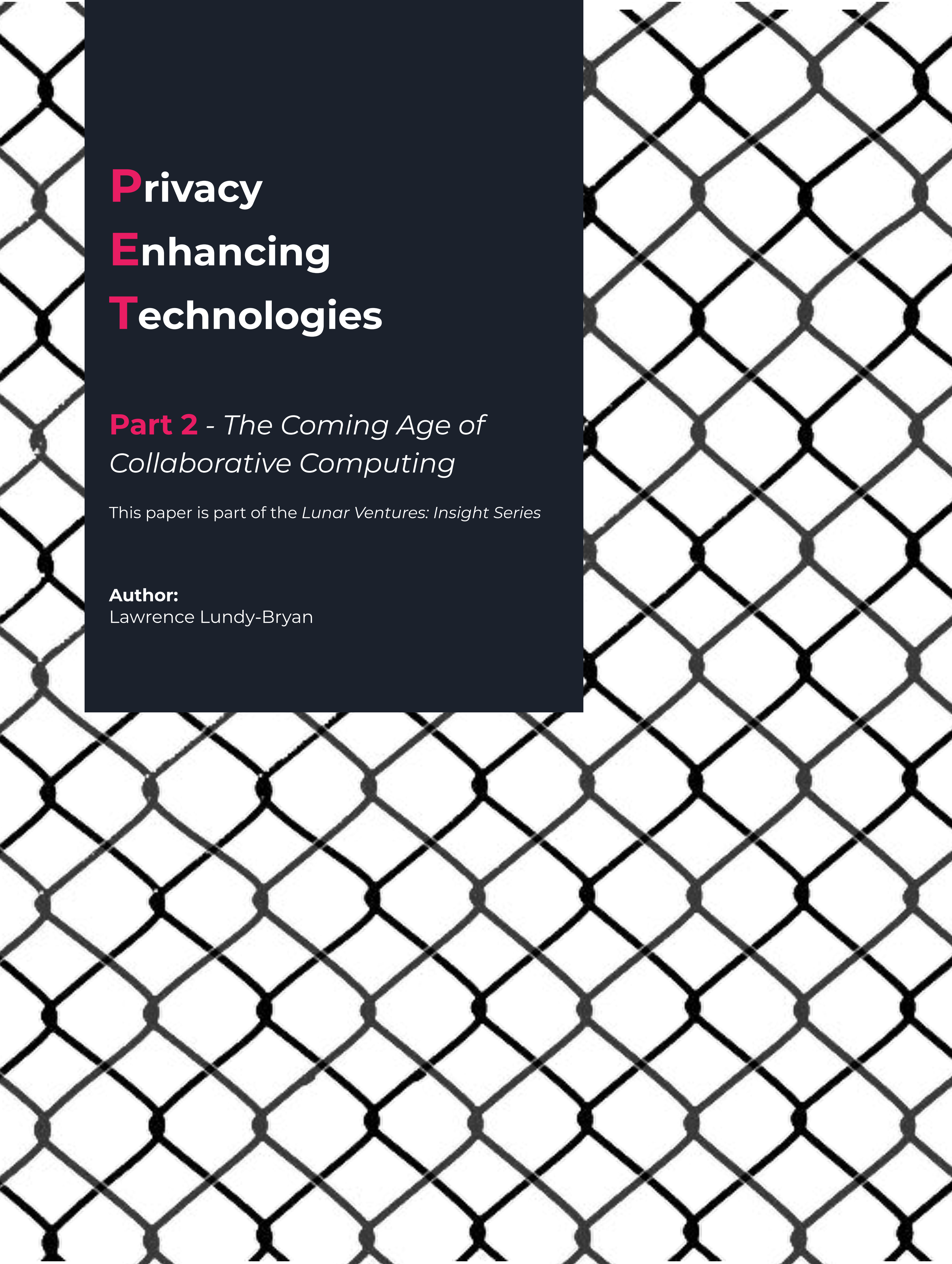


Privacy Enhancing Technologies

Part 2 - *The Coming Age of Collaborative Computing*

This paper is part of the *Lunar Ventures: Insight Series*

Author:
Lawrence Lundy-Bryan

Disclaimer

This document (the "Document") has been prepared by Berlin Innovation Ventures GmbH ("Lunar Ventures"). Lunar Ventures is registered in Germany at the Local Court of Charlottenburg with registration number HRB 190056 B. Lunar Ventures is an appointed representative of Berlin Innovation Ventures 1 GmbH & Co. KG ("Lunar I") which is authorized and regulated by the German Federal Financial Supervisory Authority.

No undertaking, warranty or other assurance is given, and none should be implied, as to, and no reliance should be placed on, the accuracy, completeness or fairness of the information or opinions contained in this Document. The information contained in the Document is not subject to completion, alteration and verification nor should it be assumed that the information in the Document will be updated. The information contained in the Document has not been verified by Lunar Ventures, Lunar I or any of its associates or affiliates. The Document should not be considered a recommendation by Lunar Ventures, Lunar I or any of its directors, officers, employees, agents or advisers in connection with any purchase of or subscription for securities. Recipients should not construe the contents of this Document as legal, tax, regulatory, financial, or accounting advice and are urged to consult with their own advisers in relation to such matters. The information contained in the Document has been prepared purely for informational purposes. In all cases persons should conduct their own investigation and analysis of the data in the Document. The information contained in the Document has not been approved by the Federal Financial Supervisory Authority. This Document does not constitute, or form part of, any offer of, or invitation to apply for, securities nor shall it, or the fact of its distribution, form the basis of or be relied upon in connection with any contract or commitment to acquire any securities. Any forecasts, opinions, estimates, and projections contained in the Document constitute the judgement of Lunar Ventures and are provided for illustrative purposes only. Such forecasts, opinions, estimates, and projections involve known and unknown risks, uncertainties and other factors which may cause the actual results, performance or achievements to be materially different from any future results, performance or achievements expressed or implied by such forecasts, opinions, estimates and projections. Accordingly, no warrant (express or implied) is or will be made or given in relation to, and (except in the case of wilful fraud) no responsibility or liability is or will be accepted by Lunar Ventures, Lunar I or any of its directors, officers, employees, agents or advisers in respect of, such forecasts, opinions, estimates and projections or their achievement or reasonableness. Recipients of the Document must determine for themselves the reliance (if any) that they should place on such forecasts, opinions, estimates and projections. Information contained in the Document may not be distributed, published or reproduced in whole or in part or disclosed to any other person. The distribution of any document provided at or in connection with the Document in jurisdictions other than Germany may be restricted by law and therefore persons into whose possession any such documents may come should inform themselves about and observe any such restrictions.

About this white paper

About Lunar Ventures

Lunar Ventures is an early seed venture fund based in Berlin. We invest anywhere in Europe in technical teams building moonshot infrastructure software companies. Get in touch with us at hello@lunarventures.eu

NEWSLETTER

About this paper

This paper is a follow-up to "*Part 1: Privacy-enhancing technologies: the privacy infrastructure of tomorrow is being built today*" (**link**).

About the author

Lawrence Lundy-Bryan, Research Partner

Technology researcher and investor trying to figure out what the world might look like in the future and invest in startups building it. Worked on investments, research and policy in the past with the World Economic Forum, OECD, RadicalxChange, and Outlier Ventures.

Contents

Acknowledgements

Thanks to the Lunar team for feedback and revisions, in particular Elad Verbin. And Andrew Trask for coming up with the phrase “partnership-enhancing technologies”.

1. Aggelos Kiayias, The University of Edinburgh
2. Alain Brenzikofer, Supercomputing Systems
3. Alex van Someren, Amadeus Capital Partners
4. Andrea Carmignani, Keyless
5. Andreas Fauler, Rocketstar Foundation
6. Andrew Trask, OpenMined
7. Archie Muirhead, IQ Capital
8. Bart Vandekerckhove, Volta Data
9. Ben Fielding, Gensyn
10. Ben Livshits, Brave
11. Can Kisagun, The Secret Foundation
12. Caroline Kaeb, EU Commission
13. Charlotte Slingsby, Nettoken
14. Claudia Diaz, Nym Technologies
15. Dan Bogdanov, Cybernetica
16. Don Gossen, Keyko
17. Fabian Eberle, Keyless
18. Francesco Gadaleta, Amethix
19. Gilbert Hill, Tapmydata
20. Graham Steele, Cryptosense
21. Harry Halpin, Nym Technologies
22. Hassan Mahmud, Digital Catapult
23. Hrishikesh Dewan, Zirohlabs
24. Irina Haivas, Atomico
25. Jason Brenier, Georgian Partners
26. Jason McFall, Privitar
27. Jason Teutsch, Truebit
28. Jonathan Rouach, QED.it
29. Jon Geater, Jitsuin
30. Jordan Brandt, Inpher
31. Jules Schwerin, RTP Global
32. Kelly Olson, Supranational
33. Kenny Paterson, ETH Zurich
34. Khaled El Emam, Replica Analytics
35. Kim Laine, Microsoft Research
36. Lauro Vanderborght, Digita
37. Marvin Tong, Phala Network
38. Maiko Meguro, EU Commission
39. Nicole Lai, Atomico
40. Nigel Smart, KU Leuven
41. Paul Francis, Max Planck Institute for Software Systems
42. Rachel O’Connell, Trust Elevate
43. Rand Hindi, Zama
44. Rick Hao, SpeedInvest
45. Ruben Verborgh, Ghent University
46. Sharon Goldberg, Boston University
47. Simonetta d'Ottaviano, Nettoken
48. Stijn Christiaens, Collibra
49. Sunny Kang, Inpher
50. Thomas Walton-Pocock, Aztec Protocol
51. Tobias Hann, Mostly.AI
52. Yehuda Lindell, UnBound

Recommendations

Founders: You're not selling privacy

1. Few people care about privacy in the enterprise. Internalise that. Build a business, not a public good.
2. Don't sell technology, sell solutions. Be clear which problems are getting solved: data liability; using personal data; outsourcing risk; collective processing; or data acquisition. And be sure to answer why this is the right balance of security, cost and performance to solve the problem.
3. Focus on the pain today. Grow from there. This is generally obvious in start-up-land, but oftentimes start-ups in this field try to solve a problem that doesn't exist yet. Founders need to reduce risk for their customers first; then solve issues of collective processing and data acquisition. As you expand your footprint in your customers' organisation, target the growing collaborative computing opportunity.

Investors: PETs are creating borderless computing infrastructure

1. PETs are a data management tool first, and a privacy tool second. Dismissing PETs because privacy isn't a large enough consumer market or GDPR compliance isn't investable is missing the opportunity.
2. The terms privacy-enhancing technology is misleading. Partnership-enhancing technologies and collaborative computing are better frames for investors, and should be attractive to any investor covering Cloud, b2b SaaS, enterprise, big data, or AI.
3. PETs are enabling data markets. A good way of viewing PETs is as a data market driver and for investors with investments in enterprise software, a portfolio value multiplier. Enabling businesses to generate value from internal data will unlock big revenue opportunities. But enabling businesses and markets to collaborate and compute on shared data is the next era in the data economy.

Policymakers: PETs support digital free trade

1. Policymakers must untangle different market failures. Problems of monopoly power require different solutions than issues resulting from individual reasoning failures from trading privacy for information goods. Natural monopolies and reasoning failures will have different policy instruments.

2. Be aware PETs offer a viable market solution to the privacy problem. Before using regulatory tools, supporting the development and commercialisation of PETs can have powerful positive externalities and drive innovation.
3. Consider PETs in the broader context of data policy and digital markets. PETs support markets and economic growth, while privacy regulation (and legal ruling like Schrems II) protects individual rights at the cost of growth. Individual rights and economic growth do not need to be at odds. PETs can square the circle: protecting individual rights while supporting trade and growth. If Government's want to support free trade, then PET-enabled digital free trade is a crucial policy area to support.

Executive Summary

Privacy-enhancing technologies (PETs) are coming of age. The tools are maturing, but mostly, it's because Gartner says so. Well no, actually it's because the market is beginning to understand what they can be used for and how that helps their business. First it's for compliance but more importantly and less well-understood it's for collaboration. Kicked off by GDPR and the wider public conversation around privacy, organisations have been forced to prioritise how they manage personal data. PETs are tools that can bring privacy features to a whole host of applications including important areas like encrypted databases and anonymous communication networks which are not covered in this paper. We explore the role PETs play in computing: specifically how they address data liability, outsourcing risk, and processing personal information, as well as opening up new opportunities for collective processing and data acquisition. We see differential privacy, synthetic data, trusted execution environments (TEEs), verifiable computation, zero-knowledge protocols, federated learning, secure multi-party computation (MPC), and homomorphic encryption as solutions to reduce risks. But some of them, especially MPC, federated learning and homomorphic encryption, open up never before possible opportunities around data collaboration. This is not the story of compliance (that would be a boring story). It is the story of how sharing beats hoarding and digital ecosystems beat monopolies. It's a good versus evil story.

PET adoption is being driven by five surprisingly non-tech trends. First, compliance legislation essentially created a market for privacy tech in the enterprise, raising the issue higher on the CTOs to-do list. Second, migration to the cloud has now reached business critical software and sensitive workloads, and with it comes risks that need more than just expensive and ineffective SLAs. Third, digital ecosystems, too, are increasing the need for multi-party coordination. Few organisations have all the necessary skills, data, and capacity to generate cutting-edge insights alone. The FAANG companies have to spend millions to do the sort of cutting-edge machine learning to drive their products. One way to solve this problem is to collectively work with partners and customers on inputs and share the outputs securely. Fourth is the fact talented people continue to develop PETs as a form of civic engagement rather than primarily for the money. For many people, especially technologists, protecting privacy is a political goal and human right. This means we see far more development than what might be expected from the size of the market. Finally, and related to civic technology, is the cryptocurrency market. This largely unregulated and dynamic market

provides an experimental breeding ground with users who are both philosophically and commercially aligned to protecting privacy. The hardest thing for a start-up is to find customers with a strong enough pain point to buy a totally new and half-finished product from a company likely to be out of business in 12 months. The crypto market despite its flaws, has plenty of these customers, perfect for a fledgling privacy start-up.

That's where the good news ends. There are lots of market restraints that we expect to slow the adoption of PETs. First, the common lack of market education and lack of talent which stalks the PET market. But that is normally the case for all new technologies and is generally overcome eventually. Second, PETs are expensive. at least compared to the alternative of processing data 'in-the-clear' without encryption. That said, PETs, at least not yet, are not competing to be cheaper or faster. The tools are superior in one dimension: hiding the inputs, operation and/or outputs of computation. Some customers consider this a huge pain point and are willing to solve it today. Over time and with investment, performance will improve so that it is no longer materially different than non-PET tools. At least for products that are closer to the application-specific end of the spectrum rather than general-purpose. Another issue, and one that is unlikely to go away with more resources and money; integration is particularly difficult because the entire development environment is different. Any integration with other software opens up a new and ongoing privacy risk which in turn makes development and maintenance expensive. This isn't the case with TEEs and synthetic data however, so we can expect these tools to find a market faster. The final problem is a lack of buyer sophistication, this is the risk that buyers don't value security highly enough to buy 'good' PETs versus software that just claims to protect privacy or be secure. The positioning of PETs as partnership tools as proposed in this paper, does make this less of a restraint by changing the buyer and the value proposition.

Looking at the drivers and restraints, we predict PETs will have a major impact on the cloud and machine learning markets. We expect TEEs to become widespread in cloud environments, but software-based cryptographic PETs will have less of a market meaning the value-chain is unlikely to change. The same is true of machine learning. The move to partnership-enhanced machine learning, enabled by the widespread use of federated learning for greater access to data, will grow the market but not upend the market. Vendors will placate the public and rile up privacy campaigners with "data doesn't leave your phone" slogans. PETs in the context of the Cloud and machine learning essentially grow the market and entrench current incumbents. Once PETs become a part of the corporate software stack, the real value can be unlocked: collaborative computing.

Expect the emergence of a larger, (maybe not global as the Splinternet becomes ever more entrenched), liquid computing and data market. We will move from cumbersome bi-lateral and multilateral data owner-data processor relationships to a more dynamic, algorithmically driven data processing and analytics market. Basically programmatic advertising exchanges but for all computing tasks. The reason not to collaborate is the fear of exposing data or confidential information. If that can be mitigated, collaboration will thrive. Owners and processors, buyers and sellers can operate in a zero-trust environment making collaboration cheaper, faster, and easier.

We predict collaborative computing to be the largest new technology market to develop in the 2020s. By 2030, data marketplaces enabled by PETs, in which individuals, corporates, machines and Government's trade data securely, will be the second largest ICT market after the Cloud.

"PETs are partnership technologies first, privacy technologies second. Let's call them partnership-enhancing technologies. Once the value proposition becomes obvious to the market, we predict that by 2030 a collaborative computing market will be one of the largest markets in the technology industry."

We recommend founders, investors and policymakers prepare:

- **Founders** need to sell increased revenues, not the technology. Be clear exactly which problems are being solved: data liability, personal data processing, risk of outsourcing to third parties, collective processing, or data acquisition, and why your answer is the right balance of security, cost and performance is the most efficient solution. If you want to tap into the bigger opportunity: sell collaboration not privacy.
- **Investors** should understand PETs in the context of borderless digital infrastructure for the Cloud and machine learning, effectively tools that can grow the market for most software today.
- **Policymakers** should think of PETs as tools to support open societies and digital free trade. More specifically these tools can be a programmatic complement for regulation offering the potential to close the gap between the speed of innovation and the speed of regulation. Using PETs to augment laws, has the advantage of protecting privacy but also encouraging innovation and economic growth.

Part 1
The Future

What are privacy-enhancing technologies and what problems do they solve? Two great questions. PETs are a class of technical measures designed to preserve the privacy of individuals or groups. We consider differential privacy, federated learning, homomorphic encryption, multi-party computation, synthetic data, trusted execution environments, verifiable computation, and zero-knowledge proofs as the main PETs[1][2]. For an in-depth look at the market today including technical readiness, application areas and use cases, read Part 1 of our PET Insight Series.

In order to make predictions, it's important not just to know the solutions, but the problems these techniques might solve. Readers familiar with the market will know PET-based products today are predominately pitched to address business risk, three problems in particular:

1. Data liability: the legal problem of collecting personal information
2. Outsourcing risk: the risk of using a third-party provider to process data
3. Processing personally identifiable information (PII): the challenge of processing datasets with PII in it whilst remaining compliance with privacy legislation

There are two more problems that are rarely seen as problems PETs can solve, but will be crucial in the adoption of PETs:

4. Collective processing: the opportunity to combine data with different parties for shared processing
5. Data acquisition: the need for more data to train machine learning algorithms

No single technique will solve all of the problems. Techniques are suited to one of more particular problem and require different design trade-offs. Nevertheless, understanding the sorts of problems that can be solved with PETs is crucial in predicting adoption and weighting the probability of different scenarios. And with that...

Welcome to the future. It's 2030 and the stock market is all meme stocks. Sorry. But Elon did get us to Mars so it was all worth it. And despite the barriers, partnership enhancing technologies have transformed computing. Privacy-enhancing technologies, so called PETs, are like privacy wrappers around cloud infrastructure and machine learning

[1] Trusted execution environments could be considered a hardware-based verifiable computation technique for our framework. However on balance, we decided it was worth pulling it out individually.

[2] The Royal Society (2019). Protecting privacy in practice: the current use, development and limits of Privacy Enhancing Technologies in data analysis. (n.d.). [online] Available at: https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/privacy-enhancing-technologies-report.pdf?la=en-GB&hash=862C5DE7C8421CD36C105CAE8F812BD0

tools. These wrappers changed the way developers built software. Just as containers made deploying software across different environments easier; PETs made deploying software safer. The assumption that that data is safe and private when shared changed the Cloud and Machine Learning markets, and in doing so, ushered in one of the largest markets in history: collaborative computing.

Application 1: Partnership-enhanced Cloud

As mentioned in the previous section, the cloud is still big business. Likely, a trillion-dollar business a year soon. Confidential cloud, or as Gartner have termed it: privacy-enhancing computation will become a major growth segment in the next few years[3]. The Confidential Computing Consortium is already using this use case to promote the use of TEEs[4]. Run by the Linux Foundation and including every large cloud provider such as Alibaba, Baidu, Google Cloud, Microsoft, and Tencent, the consortium all views this use case as a business driver throughout the 2020s. Even if the consortium fails in its objectives, its existence shows how important these companies see confidential computing in the Cloud.

There are plenty of security concerns which have cast doubt on the security guarantees of TEEs. So, we expect highly sensitive tasks to combine TEE execution with other PET tools such as homomorphic encryption and MPC for additional guarantees, especially in the security and defence industries. Software-based cryptography like verified computing may well be cheaper and certainly more flexible in the long-term, but hardware-based TEEs are good enough today. As virtualized TEE execution gets easier and more widely offered, it will be harder for software-only solutions to win customers despite having better security credentials. TEEs have the advantage over techniques like FHE and MPC that also address the untrusted outsourcing problem in that they are application specific. Simply, the product is an easier sell. Ultimately, because of path dependency or institutional inertia, good enough today wins over better tomorrow. The outstanding question regarding confidential cloud computing is how much processing needs to be confidential? The answer may well be a function of cost; how quickly competition and economies of scale bring down TEE costs may well drive the growth of this segment. The fact that all the biggest technology companies are involved in the space suggests the future for TEEs and confidential cloud is bright.

2030 Prediction

- Partnership-enhanced cloud is the largest segments of the Cloud market, with all vendors offering TEE-based environments. The TEE market expanded considerably, as hardware costs continued to be driven downwards, almost all

[3] Gartner (2020). Gartner Top Strategic Technology Trends for 2021. [online] Available at: https://www.gartner.com/smarterwithgartner/gartner-top-strategic-technology-trends-for-2021/.

[4] Confidential Computing Consortium Defining and Enabling Confidential Computing. (n.d.). [online] Available at: https://confidentialcomputing.io/wp-content/uploads/sites/85/2019/12/CCC_Overview.pdf

devices operating at the edge by 2030 have a TEE.

- Software-based cryptography such as MPC, FHE, and zero-knowledge, despite cost and performance improvements in the 2020s, are used sparingly for the vast majority of cloud environments as sunk costs on TEE workflows reduced the demand for incremental improvements. Customers who are happy to pay for a 10x security improvement along some dimension like Government and defence, will be avid consumers of FHE. Cybersecurity became a major national security concern after almost daily ransomware attacks and hacks, communication networks and other critical infrastructure invested heavily in cryptography as one layer of defence against adversaries.
- PETs won't change the cloud value chain. Buyers will continue to select cloud vendors based on cost, convenience and security. And with the major players already offering TEEs to improve their security credentials, it's difficult to see new entrants competing on cost or convenience. Disruption, if it comes, will likely come after 2030 with a shift from a client-server model to a peer-to-peer model. The seeds of this ultimate disruption can already be seen in a nascent stage in the cryptocurrency market, especially experiments with peer-to-peer markets from computing resources.

Application 2: Partnership-enhanced Machine Learning

Data quality and availability (and computational cost at the cutting edge[5]) will increasingly restrain machine learning development at the cutting-edge. Few companies can find new ways to acquire ever more data, especially as the political environment turns frosty to unconstrained data collection[6]. Google, with Federated Learning of Cohorts (FLoC) and others are already innovating in the federated learning space to retain access to data at the edge. Not only does federated learning provide new data for algorithms, but it also has the PR benefit

[5] www.yuzeh.com. (n.d.). How much did AlphaGo Zero cost? [online] Available at: https://www.yuzeh.com/data/agz-cost.html.

[6] Al-Rubaie, M. and Chang, J.M. (2019). Privacy-Preserving Machine Learning: Threats and Solutions. IEEE Security & Privacy, 17(2), pp.49–58.

of 'protecting privacy'.[7] Cybersecurity risks are reduced by not moving local data over the Internet and aggregating it all on a central server. The objective of machine learning isn't to collect data, it's to make accurate predictions. Collecting data is a biproduct. It's not only federated learning offering the potential for better machine learning while protecting privacy. Homomorphic encryption, if configured correctly, analyses ciphertext and makes accurate predictions. Local differential privacy limits the personal data aggregated in central repositories. These processes are likely to be combined over the next few years, avoiding the need to decrypt or move data from edge devices. Adding MPC could further reduce the risk of using third-party processors by splitting the model output across different parties. Regardless of the particular combination of tools, the use of privacy-preserving machine learning tools will grow because it allows for the analysis of more data from more places across devices and parties.

Bringing any technique close to cost and performance parity with machine learning on plaintext will mean vast investment. But considering the rewards, it's already happening. In an effort to head off serious antitrust litigation and consumer backlash, Internet platforms like Google and Facebook are investing in tools to yes, protect privacy but really to protect their business model. Expect to see federated learning as the weapon of choice for incumbents.

2030 Prediction

- Widespread use of PETs, specifically (and from a taxonomy perspective, controversially) federated learning, by the major Internet companies broadly placates political and consumer privacy concerns. Expect plenty of “Giving you your data back” and “Nothing leaves your smart glasses” slogans.
- Federated learning became the dominant tool for privacy-preserving machine learning and machine learning at the edge, enabling the collection of more data without changing the underlying data value chain, and without addressing the larger issue of control and governance of trained models. When it comes to machine learning in the cloud, FHE has become a de facto standard for many types of secure inference especially highly regulated industries such as advertising, healthcare, financial and defence industries where in the cost of security breaches is now extremely high.
- The machine learning value chain is basically the same as in 2020 with the early leaders in machine

[7] Although it seems that for FLoC in particular, the market has seen through these claims and rejected the particular Google approach.

learning, maintaining their advantage. Disruption may still come in the late 2020s or early 2030s from legislative action that requires collaborative ownership or at collaborative governance of machine learning models used for critical infrastructure. Any market-led change in the value chain will come from a change in leading-edge AI approaches away from data-hungry techniques like deep learning to things like one-shot learning, reinforcement learning and more Gaussian processes[8]

Application 3: Collaborative Computing

Unlike the Cloud and machine learning which are already established markets and improved by PETs, data collaboration is a use case fully enabled by PETs. This new market is enabled by solving a number of problems including outsourcing risk, personal data processing, and data liability. Once these problems are addressed to a reasonably high level, sending datasets for processing to any geography, partner or competitor will be second nature. Some companies and teams already engage in limited collaboration, but the data owner has no technical guarantee of privacy instead relying on legal agreements and trusting service providers. Corporate data management is, to a large extent, limited and bounded by compliance and SLAs. The risk associated with data leakage is too significant to invest resources in innovating in the data collaboration space, especially if PII is involved. So, there is certainly a load of data sitting unused.

According to estimates, connecting data across institutional and geographic boundaries could create roughly $3 trillion annually in economic value by 2020[9]. Once data can be processed by anyone, anywhere in the world, security moves down to the protocol level rather than the legal or regulatory level. That is the future. Today, the challenge for vendors using PETs as a differentiating feature is buyers want the fastest, cheapest and easiest-to-use data collaboration software. PETs add cost and complexity. Data analysts are used to working with and querying raw plaintext data. Today's mental model is to bring all data in one place and then figure out what questions to ask later.

[8] Harvard Business Review. (2019). *The Future of AI Will Be About Less Data, Not More*. [online] Available at: https://hbr.org/2019/01/the-future-of-ai-will-be-about-less-data-not-more.

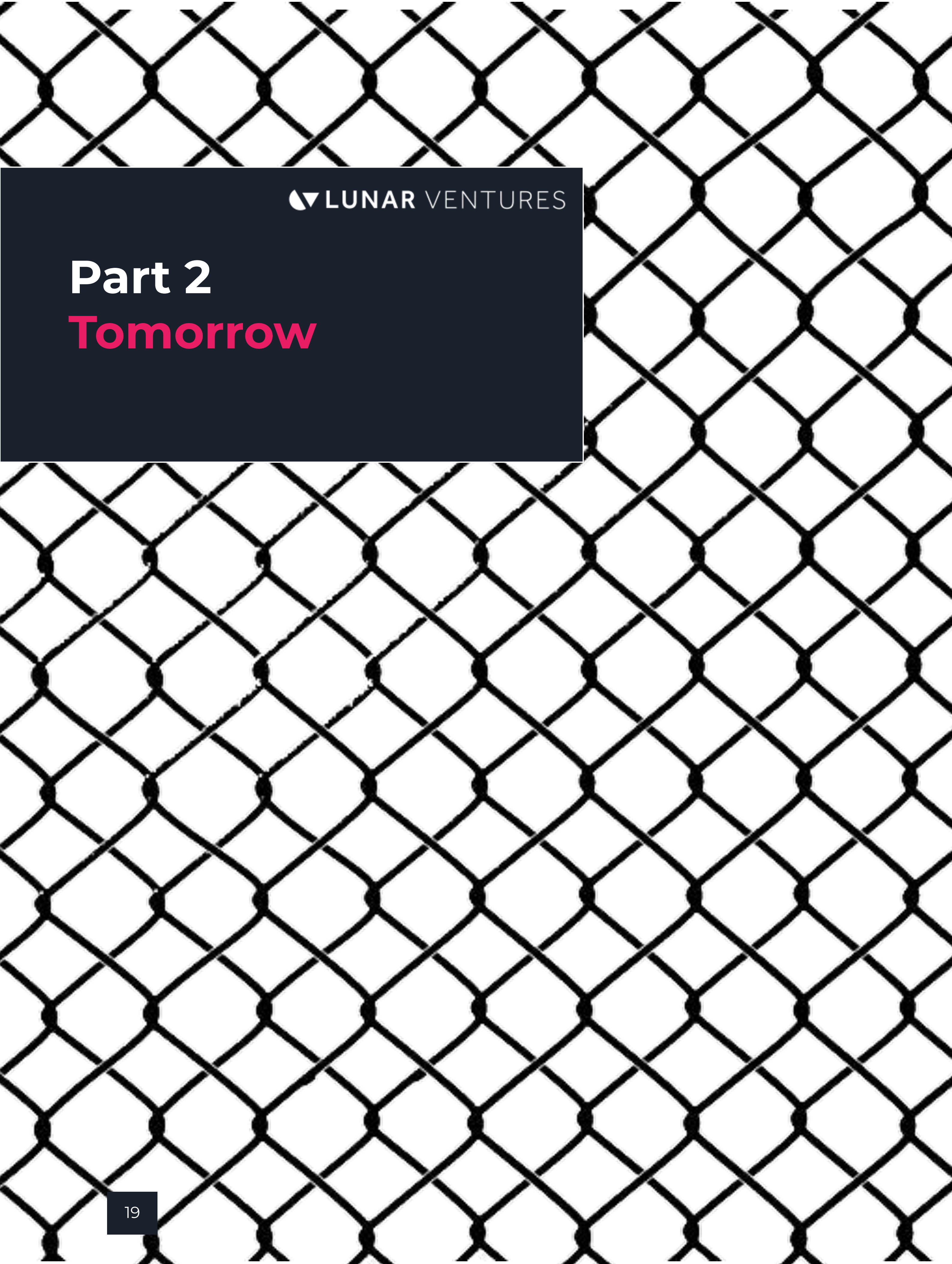
[9] McKinsey (2019a). Collaborating for the common good: Navigating public-private data partnerships | McKinsey. [online] Available at: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/collaborating-for-the-common-good#:~:text=Overall%2C%20McKinsey%20estimates%20that%20connecting.

Collaborative computing is the market that will emerge once PETs are widely used.

2030 Prediction

- Collaborative computing is the largest new market as organisations calculate the benefits outweighing the adoption costs. The market emerges from the deployment of PETs to reduce business risk in the mid 2020s.
- As an unserved market, early movers in the market will benefit substantially. Internal data collaboration will be the most straightforward pitch in the short term. Once deployed, the benefits of PETs mean that corporate or geographic boundaries are no longer a barrier. Homomorphic encryption, in particular FHE for specific use cases, will get close enough to performance parity to processing in-the-clear that performance trade-offs are less of a barrier. Zero-knowledge proofs are cheap enough to be used widely to verify computing tasks, in particular, in IoT environments, enabling machine-to-machine interactions and trading to occur with limited human oversight.
- By 2030, collaborative computing has merged two distinct segments from the early 2020s: internal data collaboration and external data collaboration. Internal data collaboration similar to Intranets of the 1990s flourished as organisations used different PETs to create workflows to connect data sources. As these tools matured and businesses, especially the legal and compliance teams, became comfortable, Intranets connected to the larger network of external data markets. Data marketplaces in which individuals, corporates, machines and Government's trade data securely could be the second largest ICT market after the Cloud.

Part 2
Tomorrow



Market Drivers

If you are wondering how we get to the wonderful world of collaborative computing, I salute your curiosity, you will be rewarded in this section.

The reality is as of 2021, the PET market is still very small, despite many of the tools being around for decades. Over the next 10 years' we expect five trends to drive adoption. We haven't included public sentiment in our forecasting. A poorly explained update to WhatsApp's terms of service in January 2021, led to privacy-focused messaging apps, Signal and Telegram to gain 7.5 million and 25 million respectively.[10] Apple continues to use privacy as a competitive differentiator for its products and services.11 And DuckDuckGo raised $100m+ in 2020.12 The last few years has seen a host of other privacy-differentiated products hit the market. These are strong signals that privacy is becoming more important to consumers. "*Privacy sceptics have dominated the discussion about online privacy for too long. Sure people care about privacy, but they'll never do anything about it. It's time to lay this bad take to rest"* as a recent blog post from DuckDuckGo put it. Despite the obvious narrative change and growth from a low base on the consider side, selling privacy to enterprise is still a challenging proposition. Privacy may have moved up slightly in the list of preferences when selecting vendors and buying software but cost, convenience and ease of use will continue to be more important. A clearer driver for corporates which is both shaping and being shaped by public sentiment is regulation, which we turn to first.

Compliance

GDPR created the enterprise privacy market, and the CCA strengthened it. Further regulation from the EU, such as the ePrivacy regulation (ePR) will continue the market growth. By 2023, data privacy regulations will cover 65% of the global population[13]. As companies invest in compliance software, they'll get a deeper understanding of their data security liabilities. Regulation won't force companies to use PETs. But as data management practices are improved, data processing becomes the weak link. A compliance value proposition should be a faster way to sell into the enterprise.

[10] The Guardian. (2021). *WhatsApp loses millions of users after terms update*. [online] Available at: https://www.theguardian.com/technology/2021/jan/24/whatsapp-loses-millions-of-users-after-terms-update.

[11] Apple (2021) Apple Privacy. [Online] Available at: https://www.apple.com/uk/privacy/

[12] TechCrunch. (n.d.). *On a growth tear, DuckDuckGo reveals it picked up $100M in secondary investment last year*. [online] Available at: https://techcrunch.com/2021/06/16/on-a-growth-tear-duckduckgo-reveals-it-picked-up-100m-in-secondary-investment-last-year/ [Accessed 16 Jun. 2021].

[13] Gartner (2020) (n.d.). Gartner Says By 2023, 65% of the World's Population Will Have Its Personal Data Covered Under Modern Privacy Regulations. [online] Available at: https://www.gartner.com/en/newsroom/press-releases/2020-09-14-gartner-says-by-2023--65--of-the-world-s-population-w

Privacy and security regulation is the Calendly link to the first zoom call. The account director can upsell over time as the procurement process is established and the vendor-buyer relationship strengthens. The risk is that a product that is pitched as a compliance tool cannot break out of that silo and tap into budgets in other more strategic business units. But having a vendor relationship with a client certainly establishes credibility which is the hardest nut to crack in enterprise sales, so on balance, with PET-based products demonstrating value is probably more important than who is the first buyer within a customer. As companies clamber to be complaint, PETs have a window of opportunity.

> ***"Quantum computing and blockchains might be cool, but do you know what's really cool? A trillion dollar a year market."***

Cloud Migration

The Cloud has been a trend in the ICT industry for the past ten years or more, but it's worth reflection. According to the Economist, the share of IT spending going to the Cloud is approaching 10%, already amounting to an annual market of $240 billion. With expected annual growth rates of nearly 20%, it could reach $1 trillion before long[14]. Quantum computing and blockchains might be cool, but do you know what's really cool? A trillion dollar a year market. And the migration of computing systems to the Cloud is nowhere near complete. Some non-core business processes and workflows have already migrated, and SaaS tools deployed, but now we are left with the hard stuff. Core business processes, commercially sensitive data, and databases of personal data. The sort of thing that poses material business and compliance risk. The solution? Stronger service-level agreements (SLAs) may go some way. But we need tools to prove correctly executed processing without decrypting the data. This gives customers a mathematical guarantee rather than a legal one. Businesses are still run on SLAs and legal contracts and we don't expect that to go away. But companies that want to reduce legal costs and liability will look to PET-based products.

Digital Ecosystems

There is an unmet need in the market for different stakeholders to work on the same data collectively. The need to collaborate on data and the tools to help do it, fit under the umbrella term digital ecosystems. There are many ways to try and solve this problem such as consortia agreements, trust frameworks and other multilateral agreements. But each of these are structured as costly legal agreements and even more costly enforcement. As good as legal

[14] The Economist. (n.d.). How Satya Nadella turned Microsoft around. [online] Available at: https://www.economist.com/briefing/2020/10/22/how-satya-nadella-turned-microsoft-around.

agreements are, they are inevitably complex and introduce overhead for start-ups and under resourced organisations operating with data. We expect software to replace much of these legal contracts over the next 10 years. As it does, digital ecosystems with programmable and cheap data sharing rules will come to dominate the market.

For digital ecosystems to reach their full potential, we will need data sharing to be simpler and baked right into the software itself. This is most obvious with data used for machine learning. Companies have collected lots of data so they can find patterns and make predictions. In the never ending search for more data, data supply chains have extended. Different individuals, teams, and organisations bring different skills, datasets, cleaning tools, frameworks, and models together to generate insights and predictions[15]. Getting all this under one roof is getting ever harder, especially when it comes to machine learning engineers. PETs enable multiple parties to collaborate on the same data without falling foul of regulators or increasing risk by leaking personal data. The environment looks more like an ecosystem of services rather than an in-house data workflow[16]. The need to work effectively with data has been key driver of an ecosystem environment[17]. But a digital ecosystem can only work when sharing is seamless and costless; meaning tools to ensure data is confidential and private.

PETs as Civic Technology

People working on PET R&D motivated by civic duty as much commercial gain. Privacy technologies occupy a unique space between private and public goods, an inherently political space. Developers inspired to build tools to protect privacy are motivated more by protecting individual freedom than becoming millionaires. This is particularly relevant because the privacy tools have continued to attract academic interest and technical development at levels not consistent with market demand. We view projects like Solid as a civic technology. Unlike the cryptocurrency market with quick riches and lambos, civic technology doesn't offer the same financial rewards. But what it lacks in money, it makes up for in reputational and social good benefits. Much like the thousands of Wikipedians doing diligent work for the community rather than for riches[18]. It’s also worth noting that many

[15] Mancuso, J. (2020). Privacy-Preserving Machine Learning 2019: A Year in Review. [online] Medium. Available at: https://medium.com/dropoutlabs/privacy-preserving-machine-learning-2019-a-year-in-review-123733e61705.

[16] Orange Business Services. (n.d.). *Growth of ecosystems in an innovation-driven world*. [online] Available at:

[17] Digital/McKinsey: Insights Winning in digital ecosystems. (2018). [online] Available at: https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Digital%20McKinsey%20Insights%20Number%203/Digital-McKinsey-Insights-Issue-3-revised.pdf

[18] Nov, O. (2007). *COMMUNICATIONS OF THE ACM*, [online] 50(11). Available at: http://faculty.poly.edu/~onov/Nov_Wikipedia_motivations.pdf.

cryptographers and entrepreneurs see privacy as their life's mission and work. More so than market opportunity, mission-driven entrepreneurs are more likely to build sustainable and successful businesses. The result is PETs have developed ahead of the market, not as a result of an identified commercial problem but to address the social challenge of widespread surveillance and the loss of individual freedoms on the Internet. This social problem and technologists search for solutions will be a core driver of PETs.

"Developers inspired to build tools to protect privacy are motivated more by protecting individual freedom than becoming millionaires".

Cryptocurrency Market

Civic tech-adjacent is the cryptocurrency market, a techno-political project is there ever was one. The cryptocurrency market is a testbed for cryptographic innovations and more specifically privacy-enhancing technologies. Many cryptographers are working in the space partly because it offers real-world implementations of their work, the potential riches probably don't hurt either. The cryptocurrency market is experimenting quickly with privacy tools like zero-knowledge, MPC, and TEEs. It makes sense because blockchains are publicly viewable, and cryptocurrency users want confidential transactions and applications using smart contracts. There is a real problem to solve, and a willing customer for those who can solve it. And not just solve it to tick a compliance box, but to really solve it in a highly adversarial environment. Cryptocurrency users are some of the only users for which security is a higher priority than usability or cost. So, the issue of buyer sophistication is less of a restraint as is it with enterprise buyers which we discuss later in the paper. The cryptocurrency market offers a first home for many PET-based products to launch and grow.

The civic tech and the cryptocurrency drivers mean PETs will mature faster than would be expected if you were to just look at commercial or consumer interest today.

Market Restraints

There are some restraints that will inevitably slow PET adoption. Some of these restraints such as education, talent, and performance are common to most, if not all, nascent technologies. Integration challenges and buyer sophistication are, however, relatively novel and potentially a serious impediment to PET deployment.

Lack of Market Education

As mentioned, buyers in enterprise settings assess PETs as privacy tools; aside from GDPR, they don't have a privacy problem. So, there is still a lot of work to be done to educate mainstream buyers on the benefits of PETs. The real challenge isn’t so much educating buyers on the ins and outs of homomorphic encryption, but more prosaically, why cybersecurity is important beyond compliance. Even if the value proposition is clear and a real need, the big challenge is the technical complexity of PETs. Organisations lack the talent to assess offerings and vendors, resulting in long sales cycles and multi-year pilots. This challenge is known as data literacy in the German Data Strategy.[19] Many start-ups waste valuable time on corporate education and risk creating consulting businesses to service client needs. We expect market education to accelerate as Google, AWS and other cloud vendors deploy TEEs and federated learning.

Lack of Talent

To develop, scale and ship PETs, a developer needs a deep background in cryptography. This problem is not uncommon with developing tools and technologies, but many PETs have been around for decades. The simple fact is that cryptography is hard, and few people do it well. And we haven't seen the required increase in the number of cryptography students in the past few years. Academic development continues to come from the same world-class teams at ETH Zurich, EPFL, KU Leuven, The University of Edinburgh and The Max Planck Institute. Unlike software engineering, cryptography is still very much an academic discipline with individuals working on cutting-edge research problems that are often quite disconnected from the market's reality. It is common for theoretical cryptography to over-serve the market on the dimension of security. Whereas usability, cost, and ease of integration are concerns rarely considered by academia. It’s not just a lack of cryptography talent holding adoption back, but the lack of product people and entrepreneurs that can understand the technical side and take cutting-edge cryptography to

[19] www.simmons-simmons.com. (n.d.). *Simmons & Simmons.* [online] Available at: https://www.simmons-simmons.com/en/publications/ckkp9wcs816rl0941llo7rbl8/new-data-strategy-of-the-german-federal-government.

market. Notable exceptions exist. But a lack of talent is a limiting factor on the number of start-ups innovating and bringing products to market.

Difficult Integration

PETs aren't just difficult to understand for the average buyer; they are challenging to use for advanced technical users, too. Most PETs, excluding synthetic data or TEEs, lack developer tools like easy-to-use libraries and frameworks. Few organisations employ cryptographers that cannot just use the tools, but also tweak them. Distributed techniques like MPC use distributed architectures rather than the traditional client-server architecture. Few developers are used to working with distributed systems and will require upskilling. Even with employees that understand PETs, a further challenge comes from integrating them with existing software and agile software development practices. Changing one part of the software stack means the privacy guarantees of the solution change; resulting in more development work to tweak the software again. Worse than that, if Google decides to change the API or library the solution uses, that means more bespoke work to ensure the privacy guarantees remain the same. Therefore, we can expect the simplest of PETs such as TEEs and synthetic data that work smoothly with existing practices to have the upper hand in terms of adoption in the short-term. In the medium to long term, widespread adoption of PETs might require changes in software development away from agile and microservices.

Poor Performance

PETs are not new. They have been around in some form or another since the 1990s. The work has primarily been on reducing the cost overhead. Working on ciphertext instead of plaintext will always suffer from some cost in performance and usability. The academic work is to reduce the costs to the point at which the trade-off for getting the benefits outweigh the costs. Some applications like fraud analytics, management of cryptographic keys, and private transactions for blockchains are ideal use cases because the performance trade-off is worth the privacy and security trade-offs. Just like complexity, the cost advantages of synthetic data and TEEs over alternatives like homomorphic encryption and zero-knowledge proofs will give them an adoption advantage.

Some applications like fraud analytics, management of cryptographic keys, and private transactions for blockchains are ideal use cases because the performance trade-off is worth the privacy and security trade-offs

Lack of Buyer Sophistication

A significant determinant of the market's growth will be to what extent the value proposition of private computing resonates with the market. A privacy or confidential computing offering doesn't operate in a vacuum; companies already use data analytics and cloud services that claim to be secure. These services have the benefit of inertia and default. The most straightforward pitch is to offer a cheaper service, although for PETs that is unlikely. So, the buyer needs to be sophisticated enough to differentiate between genuinely secure offerings and alternatives on the security front. And on the data analytics front, the buyer needs to understand and value the opportunity to process data with PII and access more data from inside and outside the company. It is therefore likely that we will see different PETs adopted in different industries based on their level of sophistication. A caveat is the less the privacy side of the value proposition is played up, and the more the data collaboration part is sold, then the less of an issue this becomes.

Part 3
Today

PET Solutions

So, there you have it. In 2030, we will be living in a world in which everybody and every machine is sharing data with abandon.

There are plenty of restraints but there are some strong headwinds driving market growth. As discussed in part 1, there are five main problems PETs address: data liability, outsourcing risk, processing PII, collective processing and data acquisition. In this last section, I want to get practical. I want to think through which techniques can be used to address these risks.

Differential Privacy

Differential privacy (DP) is a system for publicly sharing information about a dataset by describing patterns while withholding information about individuals in the dataset. An algorithm is differentially private if by looking at the output, one cannot tell whether any individual's data was included in the original dataset[20].

As an anonymisation technique, it is primarily useful for processing personal data. Local differential privacy, that is, each user adds noise to their dataset before the central aggregation stage, e.g. on a smartphone not a server, partially reduces data liability and outsourcing risk as the data processor does not hold the raw data. Global differential privacy, in which the dataset is injected with noise during the aggregation phase, does less to reduce the data liability or outsourcing risk because the processor can access the raw data. The choice between local and global differential privacy should also consider accuracy because since each user adds noise to their own data, the total noise is much larger in local models, so you need many more users to get useful results[21]. Differential privacy is limited in its application by the fact each query on the data results in some privacy loss (this varies according to the query and the calculation) and a decision must be taken to bound the total privacy loss at an acceptable level. This means differential privacy loses its value the more times the dataset is used, reducing the use cases for which it is useful. Differential privacy unlike HE, MPC and TEE, which secure the computation, secures the output of the computation. This makes it likely that differential privacy is combined with these other techniques in end-to-end privacy solutions.

[20] Harvard.edu. (2016). *Differential Privacy.* [online] Available at: https://privacytools.seas.harvard.edu/differential-privacy.

[21] Near, J., Darais, D. (2020). Threat Models for Differential Privacy. [online] NIST. Available at: https://www.nist.gov/blogs/cybersecurity-insights/threat-models-differential-privacy

Federated Learning

Federated learning (FL) is a machine learning technique to train an algorithm across multiple distributed datasets without the need to exchange them.

Federated learning isn't really designed to protect privacy, rather it's a by-product. It can address issues of data liability and the processing of personal data[22] by just not collecting the data in a central place. Fewer data are moved from local devices and aggregated, resulting in less cybersecurity risk on the processor side. Doing the learning on the device and only sending back a trained model also removes the risk of collecting, storing and analysing personal data. However, the main reason for using federated learning is to access more data to train a machine learning model. FL has the added advantage of being able to collect more data because it doesn't ever need to move the data from the local device or server. And this "the data never leaves your device" proposition makes data owners more comfortable in sharing their data. FL, like most other PETs, is rarely deployed on its own. Production systems generally combine with other tools like differential privacy and MPC.[23] There are concerns however that FL only half solves privacy concerns because it doesn't address the data outsourcing problem. This may be a fair criticism in isolation; however, the trend is to combine FL with other PETs in production to provide robust privacy guarantees.

Homomorphic Encryption (HE)

Homomorphic encryption is the property of some encryption schemes making it possible to compute encrypted data without deciphering it. Fully homomorphic encryption (FHE) supports both additions and multiplications, with partial homomorphic encryption (PHE) supporting only additions or only multiplications; and somewhat homomorphic encryption (SHE) only a limited number of both additions and multiplications.

HE has the potential to address many of the risk problems and offer the opportunity to pool and analyse shared datasets. If HE reaches performance parity with processing data in plaintext, data owners can encrypt data at source and process data anywhere and with anyone. HE is already reaching performance parity without any special hardware acceleration for targeted use cases such as the recent password breach feature from Microsoft[24].

[22] Kairouz, P., et al. (2019). Advances and Open Problems in Federated Learning. arXiv:1912.04977 [cs, stat]. [online] Available at: https://arxiv.org/abs/1912.04977

[23] Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., Kiddon, C., Konečný, J., Mazzocchi, S., Mcmahan, H., Overveldt, T., Petrou, D., Ramage, D. and Roselander, J. (n.d.). TOWARDS FEDERATED LEARNING AT SCALE: SYSTEM DESIGN. [online] . Available at: https://arxiv.org/pdf/1902.01046.pdf.

[24] Microsoft Research. (2021). *Password Monitor: Safeguarding passwords in Microsoft Edge*. [online] Available at: https://www.microsoft.com/en-

Performance parity with traditional computing is unlikely in general, but we can expect more deployments of FHE for specific applications. Unlike federated learning, data acquisition isn't the primary goal of HE. Still, by never decrypting the underlying data, it is not a leap to imagine increased data owners' willingness to share data for processing. HE has the potential to remove the requirement to trust data processors, with the obvious caveat that the processor should not be able to see the processing output, and so even HE must be combined with output obfuscation tools like DP.

Multi-party Computation (MPC)

MPC is a set of protocols allowing distributed computation on combined data without the different parties revealing any private inputs. MPC protocols include private information retrieval (PIR) allowing users to query a database whilst hiding the identity of the data retrieved and private set intersection (PSI) where two parties compare datasets to find commonalities without revealing the data to each other.

The benefits of MPC are the ability to split up and combine data and processing tasks limiting any party's control. From an outsourcing perspective, splitting up computing tasks across multi-parties is more secure than sending it all to one processor. From a collective processing perspective, these protocols allow multi-parties to pool inputs or share an output without revealing each other's inputs. The first practical application of MPC was back in 2008[25] and we are yet to see widespread use of the technology outside of key management solutions such as Unbound and Sepior. Much of this is due to the complexity of the tooling, but some of it will also be because organisations are unwilling to experiment with distributed technologies and the new workflows and development practices that they require.

Synthetic Data

Synthetic data generation is a set of artificial data derived from a source set. Generating the synthetic data preserves the real data's characteristics while protecting the personal or sensitive information present in the original data.

Exactly how close the synthetic data preserves the original data's characteristics depends on the design on the solution[26]. Still, it is a good enough solution for many businesses that want to reduce data liability and analyse the personal data they have stored but can’t currently analyse. Note, synthetic data

us/research/blog/password-monitor-safeguarding-passwords-in-microsoft-edge/

[25] Bogetoft, P., Lund, D., Damgård, I., Geisler, M., Jakobsen, T., Krøigaard, M., Dam Nielsen, J., Buus Nielsen, J., Nielsen, K., Pagter, J., Schwartzbach, M. and Toft, T. (n.d.). Secure Multiparty Computation Goes Live. [online] Available at: https://eprint.iacr.org/2008/068.pdf

[26] www.statice.ai. (n.d.). What is privacy-preserving synthetic data? - Statice. [online] Available at: https://www.statice.ai/post/what-is-synthetic-data-introduction [Accessed 8 Jun. 2021].

can as a consequence of helping clean datasets reduce data liability, but it doesn't directly address data liability. The data is still collected with synthetic data and an organisation still must be a custodian of it.

Trusted Execution Environment (TEE)

A TEE is the isolated part of secure processors that allow the isolation of secret code from the rest of the software running on a system to achieve confidentiality of the data.

TEEs aim to secure the outsourcing of program execution. Instead of trusting the third-party to perform a computation, a user uses a TEE that provides additional confidentiality guarantees. Security researchers have raised concerns around the limits of those guarantees[27]; however, the solution looks to be good enough for most companies. 95% of users will make security trade-offs against cost and usability, and side channel attack risk[28] will be, in most cases outside of defence, a trade-off worth making. It's true that TEEs don't protect against data sovereign concerns, and the limits of the TEE market will be bound by the extent to which data sovereignty becomes a predominant concern. TEEs are not the be all and end all for secure and private computing, but they solve one problem well, and that fact makes selling it that much easier. One particular issue to note however is the tendency to perform ever more processing inside the TEE itself which increases the security risk. The market will find a balance, likely stripping back the operations that actually take place inside TEEs to the minimum, which in turn reduces the number of workloads suitable for and therefore market size for TEE processing.

Verifiable Computing

Verifiable computing enables a computer to offload the computation of a function, to other perhaps untrusted clients, while maintaining verifiable results.

A useful way to think about this is as verifiable outsourcing. TEEs can be considered a type of verifiable computing that requires specific hardware. An ideal verifiable computing environment would be one completely software-based that doesn't require specific hardware. The most likely route to practice software-based verifiable computing is using FHE or probabilistically checkable proofs, but

[27] Cerdeira, D., Santos, N., Fonseca, P. and Pinto, S. (n.d.). SoK: Understanding the Prevailing Security Vulnerabilities in TrustZone-assisted TEE Systems. [online] Available at: https://www.cs.purdue.edu/homes/pfonseca/papers/sp2020-tees.pdf

[28] Bukasa, S., Lashermes, R., Le Bouder, H., Lanet, J.-L. and Legay, A. (n.d.). How TrustZone could be bypassed: Side-Channel Attacks on a modern System-on-Chip. [online] Available at: https://ronan.lashermes.0nline.fr/papers/WISTP2017.pdf

today, constructions are too costly[29]. Looking ahead 10 years, the challenge for software-only verifiable computing is to become orders of magnitude cheaper, faster or easier to use than already widely used TEE-based alternatives.

Zero-knowledge Protocols (ZK)

Zero-knowledge protocols are methods by which one party can prove to another party that they know value x, without revealing any information apart from the fact that the statement is true. For example, a prover can prove to a verifier that they graduated in 2009 without disclosing which university.

It can be argued that zero-knowledge is a subset of verifiable computing but adds the condition that the underlying data was private. Zero-knowledge protocols essentially allow transactions to occur with the disclosure of the minimum amount of personal data possible. ZK is the embodiment of data minimization, a stark contrast to transactions today where the verifier receives an abundance of personal information to verify only one or two details. This technique fits within the border privacy-by-design and data minimisation trends advocated by consumer privacy groups and regulators[30]. These techniques are in their infancy, but when deployed, businesses can substantially reduce their data liability. Zero knowledge can also be combined with a HE or MPC engine to prove that that engine did what was asked of it. In this sense, zero-knowledge can be thought of as a proof tool, something that can be added to a whole host of software to verify the correctness of the computations. The trade-off today is that it's expensive and time-consuming to generate these proofs, at least in large-scale systems. However, there is a lot of work going on to accelerate the proof generation both in hardware and software, led in particular by the cryptocurrency market.

[29] Ligeti, D., Dr, E., Peter and Barbara, M. (2019). *EÖTVÖS LORÁND UNIVERSITY FACULTY OF INFORMATICS PROOF OF ALL Verifiable Computation in a Nutshell*. [online] Available at: https://arxiv.org/pdf/1908.02327.pdf

[30] Ico.org.uk. (2019). *Principle (c): Data minimisation*. [online] Available at: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/data-minimisation/

Collective processing and data acquisition FTW

Of the five risks: data liability, outsourcing risk, processing PII, collective processing and data acquisition, the first three aren't transformational problems. They help a company be compliant with the law, improve security, and maybe act as a competitive differentiator for the privacy-conscious customer segment. But it's collective processing and data acquisition that are gamechangers.

By making it easier to confidentially pool data for processing, smaller companies lacking in machine learning capacity can collectively train large deep learning models and benefit from the trained model outputs. This reduces the costs of doing cutting-edge machine learning and will drive even faster progress in AI deployment. Adjacently, tools that make it easier to train models on distributed datasets open up new sources of training data supply for deep learning algorithms. MPC and federated learning are the most likely candidates to solve these problems. Homomorphic encryption is not explicitly designed to make collective processing or data acquisition easier, but address these challenges as a by-product of enabling processing on encrypted data.

> ***"By making it easier to confidentially pool data for processing, smaller companies lacking in machine learning capacity can collectively train large deep learning models and collectively benefit from the trained model outputs"***

Conclusion

Privacy-enhancing technologies will have a profound impact on the technology industry. Not only will they support the emergence of partnership-enhanced cloud and partnership-enhanced machine learning, transforming two of the most important technology trends of this century. They will enable the growth of the largest new market in the 2020s: collaborative computing. A computing paradigm in which data can be sent anywhere for confidential and collective processing. This environment will allow individual, corporate and machine data to be traded like just stocks on global marketplaces opening up new business models and disrupting the computing value chain in the same way the Internet has.

Big claims require big evidence, and I have walked through the inevitability of this scenario. Collaborative computing is the logical consequence of the widespread deployment of privacy-enhancing technologies. PETs deployment will be driven by five major market drivers: compliance, cloud migration, digital ecosystems, civic technology, and the cryptocurrency market. The market will grow despite strong market restraints, like the lack of market education, lack of talent, difficult integration, poor performance, and lack of buyer sophistication.

But people don't buy things because of trends. They have problems that need to fix. So, tangibly when it comes to adoption today: PETs are seen mainly as cryptographic tools to reduce risk, specifically data liability, outsourcing risk, processing PII. But as outlined they do far more. In particular, they enable two really important processes: collective processing and data acquisition. Today it is not well understood that PETs can address these processes, and we propose a term to help the market grok the opportunities: partnership-enhancing technologies. When collective processing and data acquisition are addressed, entirely new data collaboration applications will be unlocked reducing the costs of data processing and machine learning, and open up avenues for new data-based business models.

To make this collaborative computing vision a reality, we have some recommendations for founders, investors and policymakers. Founders need to sell increased revenues, not the technology. Be clear exactly which problems are being solved: data liability, personal data processing, risk of outsourcing to third parties, collective processing, or data acquisition, and why your answer is the right balance of security, cost and performance is the most efficient solution. If you want to tap into the bigger opportunity: sell collaboration not privacy. Investors should understand PETs in the context of borderless digital infrastructure for the Cloud and machine learning, effectively tools that can grow the market for most software today. And policymakers should think of PETs as tools to support open societies and digital free trade. More specifically these tools can be a programmatic complement for regulation offering the potential to close the gap between the speed of innovation and the speed of regulation. Using PETs to augment laws, has the advantage of protecting privacy but also encouraging innovation and economic growth.

"Collaborative computing enabled by partnership-enhancing technologies is the future of computing. It doesn't matter if you are shaping the future with code, money or policy, get in touch to shape it together."

LUNAR
VENTURES