Federated learning: an introduction

Federated learning: an introduction

Considerations and practical guidance for prospective adopters

Open Data Institute

Contents

About

This report was researched and produced by the Open Data Institute (ODI), and published in January 2023. Its lead authors were Anastasia Shteyn, Konrad Kollnig and Calum Inverarity. If you want to share feedback by email or would like to get in touch, contact the federated learning project lead Calum Inverarity at calum.inverarity@theodi.org.

Executive summary

After decades of increased data collection, sharing and use that has driven the emergence and development of new industries, public sentiment has been trending towards the demand for greater data privacy.

At the same time, concerns around data privacy, commercial sensitivity and security have contributed towards hesitation and reluctance to share data that might otherwise deliver significant social, economic and environmental benefits.

Privacy enhancing technologies (PETs) present potential means to facilitate greater sharing of sensitive data and to protect individuals' dignity, autonomy and fundamental rights, including data protection and privacy. Federated learning is one technology that is approaching a stage of relative maturity, in terms of awareness and practical application. It can be used to train machine learning (ML) models in a distributed manner, whilst keeping raw sensitive data safe in its original locations.

This report is the culmination of research undertaken by the Open Data Institute (ODI) between April 2022 and January 2023, supported by the Rockefeller Foundation.

In this report, we provide a comprehensive account of federated learning. We cover its primary distinguishing characteristics and the promise that it holds both for commercial use cases (within a single company or for collaboration across multiple companies) and organisations interested in using federated learning for public, charitable and educational purposes.

Though federated learning has shown considerable promise in areas such as health research, finance, Industrial Internet of Things (IIoT) and consumer applications, there remain relatively few examples of end-to-end implementations of federated learning to date. Many pilot projects and initiatives are ongoing.

Through our research, we found that privacy and confidentiality are not the most compelling benefits of this technology when deployed in isolation – that is, without additional privacy measures. Instead, the primary drivers for federated learning adoption are often scalability, improved resource utilisation and model performance improvements.

Later in the report, we consider four key dimensions that may help determine the complexity and rigour of federated learning governance, including the number of organisations involved, the level of trust between them, the design of the federated architecture, and data sensitivity.

The final section is dedicated to practical guidance for organisations in the form of a summary of proposed steps for approaching, experimenting and deploying federated learning.

Introduction

Why federated learning?

Data stewardship is the foundational activity in the lifecycle of data – collecting, maintaining and sharing it.[1] Organisations that steward data make important decisions about who has access to it, for what purposes and to whose benefit. How data is stewarded ultimately affects what types of products, services and insights it can be used to create, what decisions it can inform and what activities it can support. Stewarding data involves realising the value and limiting the harm that data can bring.

According to our theory of change, 'data hoarding' relates to a scenario where organisations restrict access to data due to misperceptions about either its value or the risks associated with data sharing. The benefits of data collection and use are only enjoyed by a few, while the negative impacts of its use affect society as a whole. 'Data fearing' describes how data might not be collected or used to its full extent, due to concerns about the harm that it can cause people being left unaddressed. People might avoid using services, or withdraw consent for data to be collected, which means that we end up missing out on data and the uses of it that could support human flourishing.

The concept of data stewardship is a tonic to these 'data hoarding' and 'data fearing' scenarios, and can help us to reach the 'farmland', where data is used to drive positive societal, environmental and economic outcomes.

Much of our work on data stewardship has focused on themes such as how decisions about data are made and the role of different types of data institutions. In this report, we explore new *technological* approaches – PETs – that can help leverage data for public, educational and charitable aims. PETs offer the potential to give access to data that may otherwise be kept closed for reasons such as privacy, commercial sensitivity or security.

Federated learning, often considered a PET, promises to train ML models in a distributed manner, keeping the raw sensitive data safe in its original locations. The fact that data doesn't need to leave its origin is the key characteristic of federated learning technologies that distinguishes it from centralised ML, which requires local datasets to be uploaded to a central server. This feature could open up new possibilities for organisations that believe that access to more data would enable better performance and accuracy of their ML models, but are unable or unwilling to share or aggregate this data.

Federated learning can be used within a single organisation to unlock data that resides in different divisions or across different devices. It can also enable ecosystems of organisations with privately held data to form partnerships that advance common goals or tackle industry-wide challenges.

Since its inception in 2016, federated learning has sparked enthusiasm in several

[1]Open Data Institute (2021) 'What are data institutions and why are they important?'

domains, such as health, finance, IIoT and consumer apps. However, we have seen relatively few successful deployments of federated learning, and those that exist are mostly in the private sector, for example by Google on Android[2] or Apple on iOS[3]. Despite some well-publicised successes, federated learning remains a rather new and underexplored area, especially in the deployment of the technology for public, charitable or educational aims.

Why does federated learning need more guidance?

Some of the promises that proponents of federated learning have made about its benefits have yet to be demonstrated in practice. The hype has yet to reach its peak, according to Gartner's 'Hype Cycle for Privacy, 2022', which suggests federated learning won't reach mainstream adoption for another 5–10 years. We set out to provide a balanced perspective to help demystify the technology and help practitioners understand the real risks and opportunities.

While there is a growing body of academic work and readily available technical tooling, relatively little practical guidance exists for organisations that may benefit from using this technology. The development and deployment of federated learning remains a relatively challenging and costly endeavour, and only some of the largest tech companies have managed to deploy it at scale. While some open-source frameworks are emerging that help organisations scale their federated learning experiments (we discuss this in further detail in the Deployment workflow section of the report), there is still a risk of knowledge lock-in – something we have already observed in other areas of ML and high-performance computing.[4] Important questions around governance, compliance and ethical aspects of federated learning remain, especially for less well-resourced organisations.

A theme that has emerged through the course of this research is the feeling that federated learning is at a critical moment. There is a risk that this technology and related PETs become the reserve of a few large companies, leading to a further consolidation of their power.[5] This risk may drive power imbalances in the data ecosystem, such as these giants becoming the de facto standard-setting authorities for federated learning, as they assume the role of infrastructure gatekeepers. Our aim is to democratise knowledge around federated learning, and diversify its use across application areas — especially in non-commercial domains where there are promising use cases but there may currently be a lack of awareness, or capacity, to adopt the technology.

Who will this guidance benefit?

We are seeking to increase awareness and understanding of federated learning — especially among actors within the ecosystem seeking to tackle important social, economic and environmental challenges.

---

[2] Google (2017), 'The Machine Intelligence Behind Gboard'
[3] Apple (2022), 'Federated Evaluation and Tuning for On-Device Personalization: System Design & Applications'
[4] Rikap, C. & Lundvall, B-Å. (2020), 'Big Tech, knowledge predation and the implications for development'
[5] Ada Lovelace Institute (2022), 'Rethinking data and rebalancing digital power'; Kelton et al. (2022), 'Virtual sovereignty? Private internet capital, digital platforms and infrastructural power in the United States'

We think this guidance will be particularly useful to:

1. **Organisations that hold sensitive data**: It may be difficult to unlock the value of data that resides in organisational silos (eg in different divisions or on customer devices). It may be unacceptable for this data to leave its origin or be processed centrally — due to its sensitive nature or compliance requirements. One of the most established use cases of federated learning concerns personalisation or customisation of consumer applications (see Google Gboard case study below).
2. **Ecosystems of organisations that hold sensitive data**: Cross-organisational data collaboration may be mutually beneficial, but it is often held back by compliance challenges or lack of trust between the organisations. Multiple organisations within one industry (eg banking) may form a consortium to tackle common industry challenges (eg fraud). Alternatively, adjacent industries (eg banking and insurance) may consider federated learning as a means of improving their corresponding business offerings. This audience may also include organisations acting to convene groups of data holders and support the deployment, such as regulators or not-for-profits that take the lead in shaping data partnerships for public, educational or charitable aims. An example is Epiverse, a project by Data.org, aiming to create data partnerships for future pandemic prevention.
3. **Policymakers, funders and other enabling actors supporting data use for social impact**: This may include governments and other organisations designing policy interventions, international policy organisations and networks, political thinktanks, and funders looking to support capacity building efforts and promote responsible governance and use of data via federated learning and other forms of PET.

In this report we address the following questions:

- What is the current status of federated learning technology, including key opportunities, real-world test cases and obstacles to adoption?
- What are the unique value drivers of federated learning that make it suitable for some use cases and less suitable for others?
- What are the key technical and organisational considerations that need to be taken into account when deploying federated learning?

Federated learning explained

What is federated learning?

Federated learning promises to train ML models in a distributed manner, keeping the raw sensitive data safe in its original locations. Like other types of ML, it is used by data scientists to solve specific problems that may involve identifying patterns, correlations or anomalies. Federated learning needs a large amount of training data, from which the model can learn in a systematic, repeatable way.

What makes federated learning different from other types of ML is that data doesn't need to leave its origin, and can be trained in its original location. To do so, the central server must communicate with the participating 'nodes', which may include devices or entire datasets. The nodes receive the model from the central server, and train it on locally held data. Then they send the updates back to the server that aggregates the changes and updates the central model. The server then provides the updated model for download to the nodes, which continue training the model and returning the updates back to the server, as often as required and as many times as necessary.

The federated learning architecture is analogous to a political federation, which is characterised by a union of partially self-governing states or regions under a central government. In federated learning, the central device – often a cloud server – serves a similar function to the central federal government of a federated state. It is necessary to mediate device-to-device or organisation-to-organisation communication. Without it, the training would be fully *decentralised* or *peer-to-peer*.

'Cross-device' vs 'Cross-silo'

Federated learning can refer to learning on data that is held on individual devices (eg phones, tablets, Internet of Things (IoT) devices) or across datasets held by different organisations or parts of the same organisational structure. This distinction is often referred to as 'cross-device' vs 'cross-silo' examples of federated learning. In the cross-device example, the number of nodes or devices can reach up to a scale of millions. In the cross-silo use case, the number of nodes is usually much smaller (eg fewer than one hundred, and sometimes only a handful).[6] It should be noted at this stage that involving fewer participants does not necessarily equate to simpler applications of federated learning and, in fact, can be more complicated, often requiring alignment of different organisations' motivations, incentives, and risk appetites.

[6] Huang et al. (2022): 'Cross-Silo Federated Learning: Challenges and Opportunities'

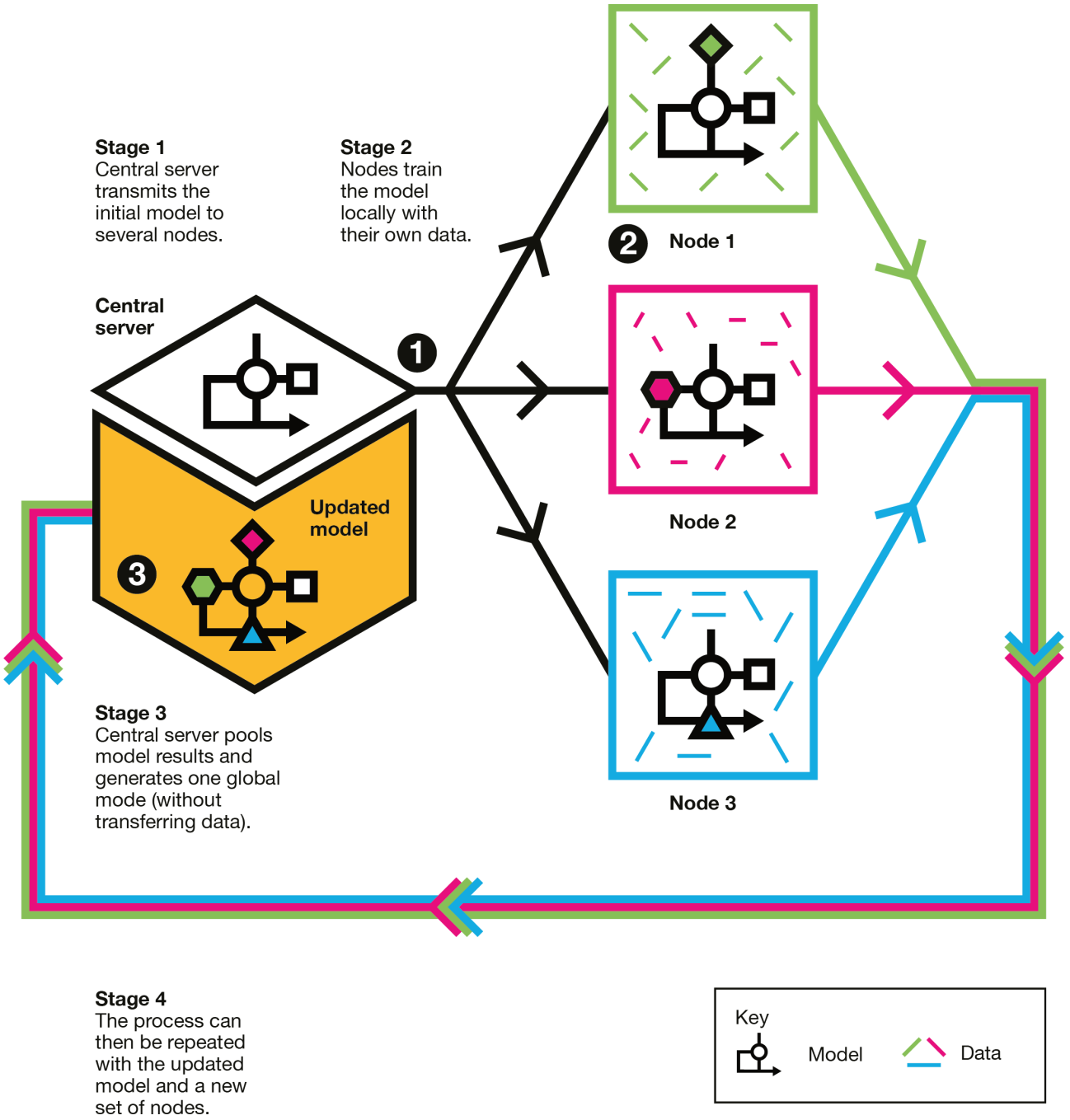


Figure 1: *Illustration of federated learning architecture and training process*

Beyond federated learning, there exists an increasing number of PETs that promise to leverage data and reduce privacy risks to individuals (such as exposing sensitive medical information when working with health data). This trend is partly motivated by the enactment of new and revised data protection and privacy laws around the globe and the increased enforcement of existing laws, such as the General Data Protection Regulation (GDPR) in the UK and EU.

Federated learning is related to other PETs – most notably Multi-Party Computation (MPC). MPC is a cryptographic protocol that enables multiple participants to run calculations on their combined data, but without revealing the underlying data to each other. Because federated learning does not – in its basic form – use encryption or other cryptographic methods to protect data, some information may get revealed at various stages of the process. It is therefore much harder to offer a strict assessment of its

security guarantees. We give a high-level overview of the PETs that are complementary to federated learning in the Common myths section of this report.

Value proposition

For organisations considering experimenting with federated learning, the value proposition tends to be two-fold:

1. Belief that access to more training data, or different types of data, can improve the ML model by making it more accurate, more generalisable, or less biased;
2. The data is sensitive, resides in silos, and/or there is a lack of stakeholder trust that prevents data sharing.
3. A single organisation may try to increase access to its siloed sensitive data. This use case often revolves around the personalisation of services, or prediction of customers' needs for marketing and advertising purposes.

Multiple organisations can undertake federated learning projects to tackle industry-wide challenges or achieve research and development objectives. Examples include using federated learning to:

- improve risk prediction in financial services,
- support more accurate and earlier diagnosis of disease, and
- enable faster volunteer recruitment for drug development in the pharmaceutical industry.

In the real world

The below table provides an illustrative summary of the most promising applications of federated learning.

It is important to note that many of the discussed applications are *speculative* or *prototypes* that have not actually been deployed in practice at scale, and may face significant challenges in trying to do so. We have also observed that cross-organisational use cases are only beginning to emerge, and it will take time until there are established social, technical and regulatory frameworks to support them.

Typology	Example application areas
Organisations that hold sensitive data that cannot leave its origin (eg due to its sensitive nature, security, or policy requirements)	Consumer apps: Federated learning can enable improvement of digital products and services by learning from device-held data. Recommender systems and online advertising are specific examples, which allow organisations to improve the overall prediction accuracy and personalisation of results. In addition to the Gboard example cited below, notable implementations include automatic speech recognition on iOS, news recommendations in the Brave browser, and privacy-preserving mobile advertising.

	This is the most mature application area in which federated learning has been deployed at scale to date.
	IIoT: Federated learning can be used to train models on condition-monitoring data gathered from IIoT platforms. Example applications include predictive maintenance of assets in manufacturing and 'smart healthcare'. Utilities providers can use federated learning to predict or optimise energy consumption in different settings (eg by using smart meter data). The challenge remains that many IIoT applications have high requirements in terms of safety and reliability, while federated learning does not usually give reliable guarantees.[7] *This is an application area that has received a lot of attention in academic work, with some pilot projects underway.*
	Mergers and acquisitions (and similar use cases): The vast amount of data gained via a merger or an acquisition can be inaccessible, or would take too long to integrate into existing storage solutions and workflows. *This is a prospective application area that organisations may wish to explore on a case-by-case basis.*
Ecosystems of organisations that hold sensitive data and are seeking to collaborate (sometimes via a convener or broker)	Health R&D: Multiple organisations, such as hospitals and research institutions, can form partnerships to securely learn from their privately held datasets, by analysing medical images and other types of data, to improve diagnosis and treatment of common diseases. In the pharmaceutical industry, recruitment of clinical trial participants can be significantly accelerated by using federated learning on siloed datasets. We are hoping that this work (among other similar projects) contributes to inspiring the use of federated learning across more of this form of collaboration, such as the Epiverse project by Data.org. We have also seen the potential of federated analytics (discussed further in the Federated analytics section) in pandemic response scenarios, specifically when applied to coronavirus (Covid-19) contact tracing initiatives.[8] *This is the most promising application area for ecosystems of organisations that hold sensitive data; it has seen several successful pilots to date. The federated analytics use case for pandemic responses is arguably more mature and established than the federated learning use case.*
	Financial services: Multiple organisations, such as banks or other financial institutions, could form consortia to develop and improve their fraud prevention/Anti-Money Laundering (AML) models. Using federated learning for credit scoring models is another example of a

[7] Boobalan (2022), 'Fusion of Federated Learning and Industrial Internet of Things: A survey'
[8] Open Data Institute (2022), 'Federated analytics for public good: Contact-tracing apps in Covid-19 containment'

	compelling use case that could help foster greater financial inclusion and create savings for individual organisations. *We identified financial services as the second most promising application area for ecosystems of organisations. Several of our interviewees have highlighted this area, although we have yet to see a substantial proof-of-concept.*
	E-government: Although we haven't seen significant successful implementations in this sector, using federated learning in an e-government context[9] could enable innovative services that could further citizen participation and incentivise private organisations to provide intelligent services in collaboration with the public sector. *Some academic sources have identified this application area. We are yet to see this deployed in practice.*

Table 1: *Typology of organisations that could benefit from the use of federated learning and illustrative applications*

Case Study: Google Gboard as the original federated learning application

The term 'federated learning' was coined by Google in 2016. It was presented as a ML approach that trains algorithms across multiple-edge devices (eg smartphones and IoT devices such as sensors), with these devices storing raw data locally and exchanging focused updates to achieve learning.

The first large-scale consumer-facing application of federated learning was Gboard, the keyboard application on Android. It was spun off by Google from the Android Open Source Project (AOSP) in 2013 and has since been developed independently. In Gboard, federated learning is used to enable functionality such as next word prediction (NWP), without sharing any user data directly with Google.

While federated learning use cases continue to emerge within industry, the Gboard use case – or later similar consumer applications such as Apple's Siri – remains a foundational example and reference point for applications of federated learning. Deployed across millions of phones, this application evidences the benefits of federated learning: creating power savings, performance improvements and keeping user data more confidential, while minimising the possibility of malicious attacks.

Google was not starting from scratch when introducing federated learning into Gboard. The company had previously gained expertise and built tooling by using federated learning to improve its internal data centres. A key challenge for Google was the lack of targeted tooling for this new area application. For Gboard, Google had to revise its existing tooling from data centre applications significantly, and introduce secure aggregation to keep users' model updates safe.

[9] Guberović et al. (2022), 'Framework for federated learning open models in e-government applications'

Figure 2. Google Gboard, the default keyboard on the Android operating system, was the first large-scale deployment of federated learning to consumers.

Over time, the success of Gboard also lessened the demand for third-party alternatives. Third-party keyboards, like SwiftKey or Swype, which used to be highly popular, have lost prominence since Google entered the keyboard market. For example, Swype, by the speech technology company Nuance, was once the default keyboard on many Samsung phones, but has now been discontinued. Interestingly, both SwiftKey and Nuance have now been acquired by Microsoft.

These observations show how large technology companies currently manage to reap the benefits of federated learning through their capital (including talent, data centres, and knowledge) and pre-existing dominance in digital markets.

Federated learning maturity and obstacles to adoption

Since its inception in 2016, a large body of academic work and industrial research effort has gone into testing the applicability of federated learning to use cases that are different from the canonical Gboard example – such as those with a small number of participants, where data is unevenly distributed across devices, or more widely across organisations and industry silos.[10]

Nowadays, federated learning has achieved considerable maturity from a technical perspective. For many applications, federated learning achieves similar performance levels to traditional ML. Free and publicly available federated learning tooling also exists that has been used for real-world deployments of federated learning. Yet, the tooling and documentation are not nearly as robust and reliable as they are for traditional ML, and require more resources to set up – with such expertise tending to be rare and costly. The distributed nature of federated learning also makes the development and deployment of federated learning applications inherently more difficult than for traditional ML. This is because code must be developed and tested across many different computers rather than just one central server.

More work is needed on the privacy and security profile of federated learning, as well as its potential to mitigate bias and unexpected outcomes. In particular, like traditional ML applications, federated learning lacks concrete privacy and security guarantees.

[10] See, for example, Xie et al. (2021), 'Multi-Center Federated Learning'

This can make it hard for organisations to rely on these technologies, and for non-experts to make trade-off decisions around the use of federated learning in practice. Due to the lack of explicit regulatory guidance and established jurisprudence (as with traditional ML), the data protection implications of federated learning can be unclear. While federated learning can allow organisations to draw on additional and potentially sensitive data (eg race, gender) to mitigate bias and unexpected outcomes, issues of bias and fairness remain with federated learning and require further consideration.

So far, it has mainly been large, commercial organisations that have managed to deploy federated learning at scale. They have the necessary infrastructure, sophisticated data centres, and the leading ML experts, to do so. The paradigm of federated learning also aligns well with their business interests: while federated learning trains ML models in a decentralised way and prevents direct data sharing, a central server is still in control of the overall operations, which may allow these organisations to benefit from the models and aggregate information generated. This also poses the risk that these few companies will develop and exert significant decision-making power over the federated learning ecosystem, such as by setting standards. The most notable attempt so far has been Google's proposed Federated Learning of Cohorts (FLoC) initiative. Google introduced the FLoC technology to make personalised online ads more privacy-preserving by grouping users into 'cohorts' on the basis of their online activities and interests. There remained significant concerns, particularly that FLoC might make online tracking even more invasive and might consolidate Google's influence in online advertising. In early 2022, Google decided to abandon FLoC, and replace it with Topics API – a similar technology that gives individuals more insights into the categories ('Topics') that may be used for tracking online advertising.

Federated learning will likely mature further in the coming years, especially when used in conjunction with other PETs. Organisations that consider using federated learning now could encounter significant challenges, but could also enjoy first-mover benefits and build unique expertise. In the near future, we should expect more ready-to-use tooling and guidance on the use of federated learning.

As discussed above, this project aims to democratise access to knowledge around federated learning, and demonstrate its potential for use cases in non-commercial sectors that may currently be hindered by a lack of awareness, or capacity to adopt the technology. In these contexts, we have seen several promising pilot projects in their pilot or proof-of-concept phase – and we are hoping to see more as the technology matures and allows for larger-scale implementations; we discuss an example in the health research case study below. Yet, practical deployment for these non-commercial use cases remain further down the road, in part because more stakeholders tend to be involved in these projects than for those led by large technology companies.

Case study: Federated learning in health research

In 2020, Moorfields Eye Hospital NHS Foundation Trust embarked on a research collaboration with Bitfount, a distributed data science platform for privacy-preserving ML and analytics. The collaboration was designed to develop and deploy ML models for the diagnosis and treatment of common eye diseases,

such as age-related macular degeneration. The platform uses federated learning to train the diagnosis models on siloed data from different locations within the trust, as well as to subsequently deploy those models in a federated privacy-preserving way. The work is supported by an Innovate UK grant as part of the UK's wider priority to become an international hub for ML.

Having previously collaborated with DeepMind, the research team at Moorfields appreciated the complexities involved in health data sharing partnerships. The initial steps here included:

- Moorfields set up the platform on its servers' private cloud and on-premises at the UCL Institute of Ophthalmology (where Moorfields' research lead, Pearse Keane, is Professor of Artificial Medical Intelligence). Notably, the set-up didn't require adding potentially dangerous exemptions to the Moorfields firewall.
- The technical set-up had to be signed off by the Moorfields Information Security team, and approved by their internal governance committee following a Data Protection Impact Assessment (DPIA).
- Initial models were trained on tabular and image data, starting with public datasets, and moving on to anonymised datasets that had been pre-approved for research purposes. As an additional precaution, only individuals who already had the appropriate permissions could access the data.
- Researchers ran preliminary experiments to estimate the duration of the federated learning training process, gauged the performance of the models, and assessed the trade-offs of introducing complementary technologies such as differential privacy. Some early research is also underway to demonstrate the effects of federated learning on fairness, as the datasets come from several locations that differ in data subjects' ethnic and socio-economic characteristics.

Bitfount focused on obtaining relevant certifications and on improving the usability of the platform (eg by making a desktop application available for both Windows and macOS).

While the collaboration is still in early development, both parties regard the partnership as strategic and see immense potential beyond the Moorfields research network. There are projects underway to collaborate with researchers across other UK centres and internationally. It may even be possible to bring the platform's capabilities to optometry practices across the country, and beyond. Federated learning platforms such as Bitfount may also open doors to collaborating with institutions that don't yet have the mature infrastructure or information governance procedures.

Federated analytics

Like federated learning, federated analytics became largely popularised as a result of efforts by Google in 2020,[11] several years after its initial work on federated learning. While this technology had previously been proposed,[12] federated analytics emerged almost as a byproduct during efforts towards developing federated learning itself. Where federated analytics differs from federated learning is that it is used to generate insights from distributed data, rather than to train models. It is worth noting, however, that federated analytics is not a wholly new technique and is closely related to existing analytics techniques like distributed databases, which have been in existence for decades.[13]

When determining whether a model trained in a federated context on decentralised data is suitable (eg to understand accuracy and bias), it can be necessary to evaluate locally held models and data. Federated analytics can perform this task, using the same infrastructure as federated learning to perform evaluation on decentralised data, rather than using this infrastructure for the purpose of training a ML model.

This approach can also be used to carry out regular analytics tasks on federated data, rather than only as a part of the federated learning process. This raises the prospect of repurposing the same federated learning architecture to conduct localised statistical analysis that could inform the development and improvement of products, while – similarly to federated learning – promising to remove the need for sharing raw data.[14] The analytical insights can include averages, heavy hitter identification[15] and more.[16]

It is important to note that many of the possible drawbacks of federated learning are also present within federated analytics. Particularly when considering possible security challenges, the present state of research on federated learning and federated analytics offers greater options to counteract or minimise potential 'inference attacks'[17], whereas 'poisoning attacks'[18] are less comprehensively covered.[19]

Federated analytics can also come with unexpected consequences for individuals' reasonable privacy expectations when data about them is analysed without their awareness and explicit consent – thereby potentially turning it into a new surveillance tool. Since data protection laws like GDPR put great emphasis on the fact that data collection must be reasonable and within the expectation of affected individuals, the use of federated analytics could still violate applicable legislation, and people's expectations, and needs to be exercised with caution.

Although federated analytics uses a similar infrastructure to federated learning, the deployed applications and research on federated analytics are not yet as extensive as those for federated learning. Through the course of our research, it became apparent

[11] Ramage, D. and Mazzochi, S. (2020), 'Federated Analytics: Collaborative Data Science without Data Collection'
[12] Florisi, P. (2017), 'Federated Analytics and the Rebirth of Data Science'
[13] See, for example, Ozsu et al. (1991), 'Distributed database systems: where are we now?'
[14] Pandey, S.R. et al. (2021), 'Edge-assisted Democratized Learning Towards Federated Analytics'
[15] Zhu, W. et al. (2020), 'Federated Heavy Hitters with Differential Privacy'
[16] Sadilek, A. et al. (2021), 'Privacy-first health research with federated learning'
[17] An inference attack is where a malicious actor uses characteristics – often of the global and local federated learning model – to deduce the data that was used for the training, which is therefore a privacy risk.
[18] A poisoning attack involves a malicious actor trying to spoil the training of the global model by contributing deliberately misleading or inaccurate data to spoil the training process.
[19] Wang, D. (2021), 'Federated Analytics: Opportunities and Challenges'

that interest in federated analytics is increasing significantly and that there is a vibrant community emerging around the use of this technology.

The possible applications, practical deployment considerations and business cases for federated analytics are presently in their relative infancy. In part, this can be attributed to outstanding privacy concerns surrounding the technology, which have inhibited deployment.[20] At the same time, a federated analytics prototype might help demonstrate the advantages of using federated technologies in a specific application area more easily than developing a prototype of an entire federated learning system.

At the time of writing, there are relatively few well-publicised examples of deployed federated analytics, beyond examples such as Google's 'Now Playing' feature,[21] which uses insights from devices to optimise regional databases for the purpose of song identification. There are, however, some promising proofs of concept in development, reflecting the growing interest in applying federated analytics. One example is the proposed use of federated analytics in conjunction with multiparty homomorphic encryption (FAHME)[22] in healthcare settings. While this proposal involves considerable computational and technical effort, it shows how federated analytics might be deployed in close alignment with GDPR.

Overall, much more work will be needed to demonstrate that federated analytics can deliver on its promises and the current hype.

[20] Froelicher, D. et al. (2021), 'Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption'
[21] Ramage, D. and Mazzochi, S. (2020): 'Federated Analytics: Collaborative Data Science without Data Collection'
[22] Froelicher, D. et al. (2021), 'Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption'

Deploying federated learning in practice

How well is federated learning suited to your needs?

This section of the report focuses on some of the practical considerations that can help organisations make an informed decision around the suitability of federated learning to their use cases, and its potential to help them realise specific benefits.

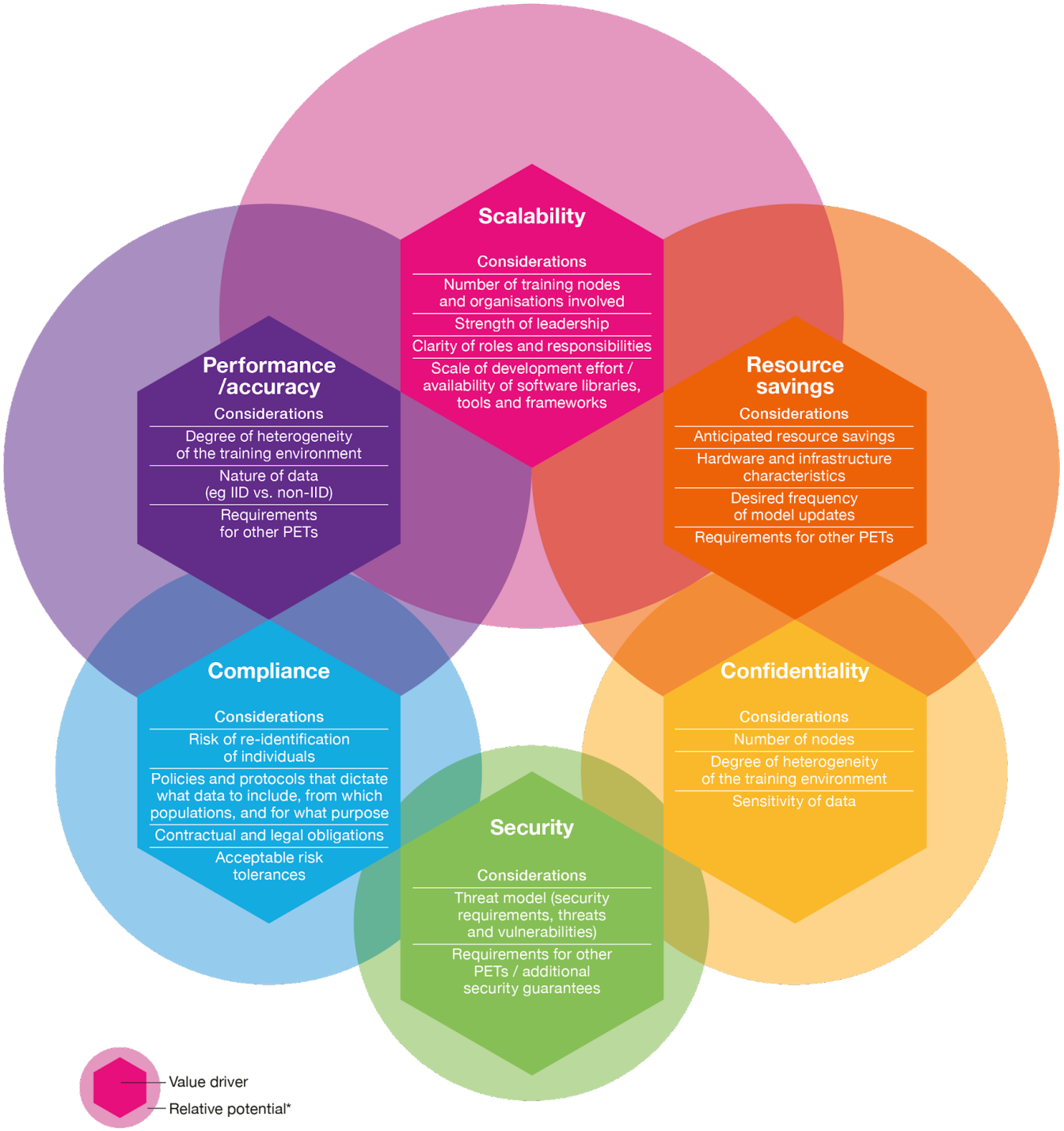
Our research has shown that while some benefits have been demonstrated in practice, others have received significant attention but are far from guaranteed. The table below provides our qualitative assessment of the potential to achieve commonly cited benefits of using federated learning.

Benefit	Benefit description	Drawbacks & trade-offs	Potential
Scalability	Federated learning could enable faster training of large data cohorts, leading to greater accuracy and time-to-value from ML Once process is established, it may be relatively easy to add additional participants	Set-up and data alignment may be effort-intensive – especially in environments with different devices, data types, and infrastructure configurations	High
Resource savings	Reduced bandwidth requirements and internet traffic Associated with lower environmental footprint than traditional ML approaches[23]	May not work with low connectivity (eg remote IoT devices with unreliable internet connection) Device owners may see reduced device performance Combining federated learning with other PETs may negate resource benefits	Medium
Model performance and accuracy	Access to more datasets/more representative data, could improve model performance May help make model suggestions more relevant and personalised	Devices may drop out before finishing training cycle due to poor network connectivity and/or power supply	Medium

[23]Qiu et al. (2020), 'Can Federated Learning Save the Planet?'

	May reduce algorithmic bias by tapping into new datasets		
Confidentiality	High volume of participants could make it more difficult to single out individuals and minimise the risk of data breaches	Model parameters that are being sent between the server and the nodes could conceal sensitive information, which can be extracted by malicious organisations or individuals	Low
Legal and regulatory compliance	Could help fulfil regulatory or contractual obligations by keeping data at source	Federated learning provides few guarantees to protect sensitive data or to ensure model outputs are ethical Poor quality of training data may result in biased model outputs Regulatory obligations impact multiple parties (eg data controllership, completion of DPIAs) Hard to quantify the risks of federated learning implementations (eg for audits) Implications for other areas of law (eg intellectual property law) are unclear	Low
Security	Combined with other PETs, it's possible to increase security of the federated learning process, eg by preventing poisoning attacks or concealing the local model updates from the server Some intermediaries/platforms can monitor outliers or potential data poisoning incidents	While security risks associated with the sharing of data may be reduced in federated learning, alternate security risks arise as model parameters are shared The two primary categories of these possible security challenges within federated learning can be considered as inference attacks and poisoning attacks (see Security section below)	Very low

Table 2: Potential benefits and associated tradeoffs for consideration when using federated learning



*The diagram above represents our qualitative assessment of the potential benefits that could be realised with FL, and may help assess the suitability of the technology to specific use cases. The size of each circle represents the relative potential to help organisations realise benefits such as scalability (highest potential), resource savings (medium potential), performance/accuracy (medium potential), compliance (low potential) and confidentiality (low potential), and security (very low potential). These benefits are interconnected and dependent on many factors; some illustrative considerations are included.

Figure 3: *Potential values to be gained through deploying federated learning, with related considerations, to facilitate assessment of suitability*

Considerations explained

- **Scalability:** The scalability potential of the federated learning environment depends on the prospective number of training nodes and, in cross-silo use cases, the number of organisations, the leadership, and the clarity of roles and responsibilities among them. It is also contingent on the availability of suitable tooling, frameworks, and software libraries, and the extent of custom development required.
- **Resource savings:** Organisations should complete an assessment of how much resource they are anticipating to save by training their ML models in a federated way. The anticipated savings may differ drastically between cross-device or cross-silo use cases, and depend on the types of devices and datasets involved, and the requirements of the training process (eg the frequency of communication between the server and the nodes). It's also important to consider the device performance trade-offs, eg for consumer devices in cross-device applications. Furthermore, our research suggests that implementation of complementary PETs alongside federated learning may negate some of the resource savings benefits.
- **Performance and accuracy:** The benefit of including additional data that may otherwise be inaccessible to the training process could be substantial. It may help make the model more representative or allow for more personalised outputs. However, the extent to which this benefit can be realised depends on the types of data and the training environment itself. There is a growing volume of academic literature on optimising the performance of federated learning in non-standard environments, for example when the data is not independent and identically distributed (Non-IID) on the local devices.[24] Combining federated learning with other PETs may result in further performance trade-offs.
- **Confidentiality:** High volumes of participants could make it more difficult to single out individuals and minimise the risk of data breaches. Another consideration here is how sensitive the raw data is. Confidentiality that can be achieved with federated learning would depend on the types of devices, data types and infrastructure configurations involved. The set-up may or may not require complementary PETs.
- **Legal and regulatory compliance:** The compliance benefits would largely depend on the risk of re-identification of the individuals within the data. Organisations with more mature data protection processes may be better positioned to structure their federated learning efforts to maximise compliance benefits (eg by enforcing strict control of what data is used for what purpose).
- **Security:** Organisations need to identify security requirements for their federated systems, assess specific security threats and potential vulnerabilities, quantify their criticality, and prioritise remediation methods. It's possible to increase security of the federated learning process by combining federated learning with other PETs.

[24] Zhu et al (2021), 'Federated learning on non-IID data: A survey'

Common myths

As demonstrated in the section above, federated learning offers some particular benefits, but there's also a lot of hype.

Legal and regulatory compliance

Federated learning is not an antidote to organisations' legal and regulatory compliance challenges. It does not offer inherent 'data protection by design'. In the context of GDPR, the federated approach brings the added complexity of identifying and sharing responsibilities across joint data controllers, completion of multiple DPIAs, and further complexity related to auditing the environments and ensuring the ML models function as expected. Although we are not ruling out some edge cases, according to our expert interviews, federated learning cannot usually be employed as a means of minimising company liability and obligations under applicable legislation (eg data protection and copyright law).

Confidentiality

The privacy and confidentiality properties of federated learning run along a spectrum from minimally (when used in isolation from other PETs) to more privacy-preserving (when combined with one or multiple PETs). Measuring the risks associated with privacy and confidentiality is often challenging, and there are no formal guarantees available at present. These challenges are shared with traditional ML approaches and are therefore not wholly unique to federated learning.[25]

Our research has found that combining federated learning with other PETs may allow for more robust privacy guarantees, but that it also comes with trade-offs. The table below summarises some relevant considerations:

Complementary PET	Benefits	Trade-offs
Differential privacy[26]	Adds 'noise' to data to make singling out individuals more difficult	Differentially private data may have less utility; context-dependent parameters can lead to unexpected data leaks if chosen wrongly
Secure aggregation[27]	Tries to hide model updates from the central federated learning aggregation server	Relies on one or more other PETs (eg MPC or TEE) that come with their own trade-offs
Secure multi-party computation (MPC)[28]	Multiple entities compute an operation on their own private data	Can be computationally expensive and difficult to scale

[25] See, for example, De Cristofaro (2021), 'A Critical Overview of Privacy in Machine Learning' and Ruehle et al. (2021), 'Privacy Preserving Machine Learning: Maintaining confidentiality and preserving trust'
[26] See McMahan et al. (2018), 'Learning Differentially Private Recurrent Language Models', for an introduction to DP in FL; for shortcomings, see, for example, Hitaj et al. (2017), 'Deep Models Under the GAN: Information Leakage from Collaborative Deep Learning'
[27] See, for a start, Bonawitz et al. (2017), 'Practical Secure Aggregation for Privacy-Preserving Machine Learning'
[28] Bonawitz et al. (2017), 'Practical secure aggregation for privacy preserving machine learning'

Trusted execution environment (TEE)[29]	The design of computing hardware restricts permitted operations on given data, and what information can be retrieved from the TEE	Requires trust in the hardware Can limit utility of data

Table 3: *A brief overview of other PETs that are sometimes considered for use together with federated learning*

For more detailed discussion, we suggest consulting The Royal Society's 'Protecting privacy in practice' report.[30]

Security

While potentially providing certain security benefits in comparison to traditional ML, federated learning concurrently gives rise to other challenges that require consideration.

A benefit of federated learning, given that raw data is not transferred, is that sensitive data is not shared and then housed elsewhere, post-training. This can be appealing, as in federated learning it is no longer necessary for the data to be stored both at its original source and training location – where another controller must be trusted to ensure the data is sufficiently protected.

However, while federated learning may reduce the security risks associated with the data-sharing, alternate security risks arise because that model's parameters are shared instead. As outlined earlier in the report, the two primary categories of these possible security challenges within federated learning are inference attacks and poisoning attacks. The former may involve reconstruction attacks and memberships attacks, where pre- and post-training models can be used to pick out characteristics of the data that was used to train the model.[31] The latter may result in model or data poisoning, which involves malicious actors intentionally trying to spoil the training through directly sending incorrect model updates or the alteration of datasets used for training.[32] These attacks look to exploit the specificities of federated learning due to the fact that it can often involve training across a wide range of distributed and potentially unreliable devices or silos, and therefore presents attack surfaces that have different characteristics to those in traditional ML settings.

Like the measures noted above that can be taken to make privacy more robust, the use of additional PETs can help when attempting to mitigate some of these attacks. Research into addressing federated learning security challenges in various architectures and environment styles is ongoing. Therefore it is vital at present for participants to consider potential security vulnerabilities before deploying federated learning and to address these as robustly as possible.

[29] See, for example, Mo et al. (2021), 'PPFL: privacy-preserving federated learning with trusted execution environments'
[30] The Royal Society (2019), 'Protecting privacy in practice: The current use, development and limits of Privacy Enhancing Technologies in data analysis'
[31] Mothukuri et al. (2021), 'A survey on security and privacy of federated learning'
[32] Kairouz et al. (2019), 'Advances and Open Problems in Federated Learning'

Governance

As is the case when deploying any novel data technology, organisations experimenting with federated learning must ensure there are strategic guardrails in place to minimise the risks and maximise the benefits intended.

In designing these guardrails for a new federated learning project, our research suggests that there are four key dimensions, or elements, to consider:

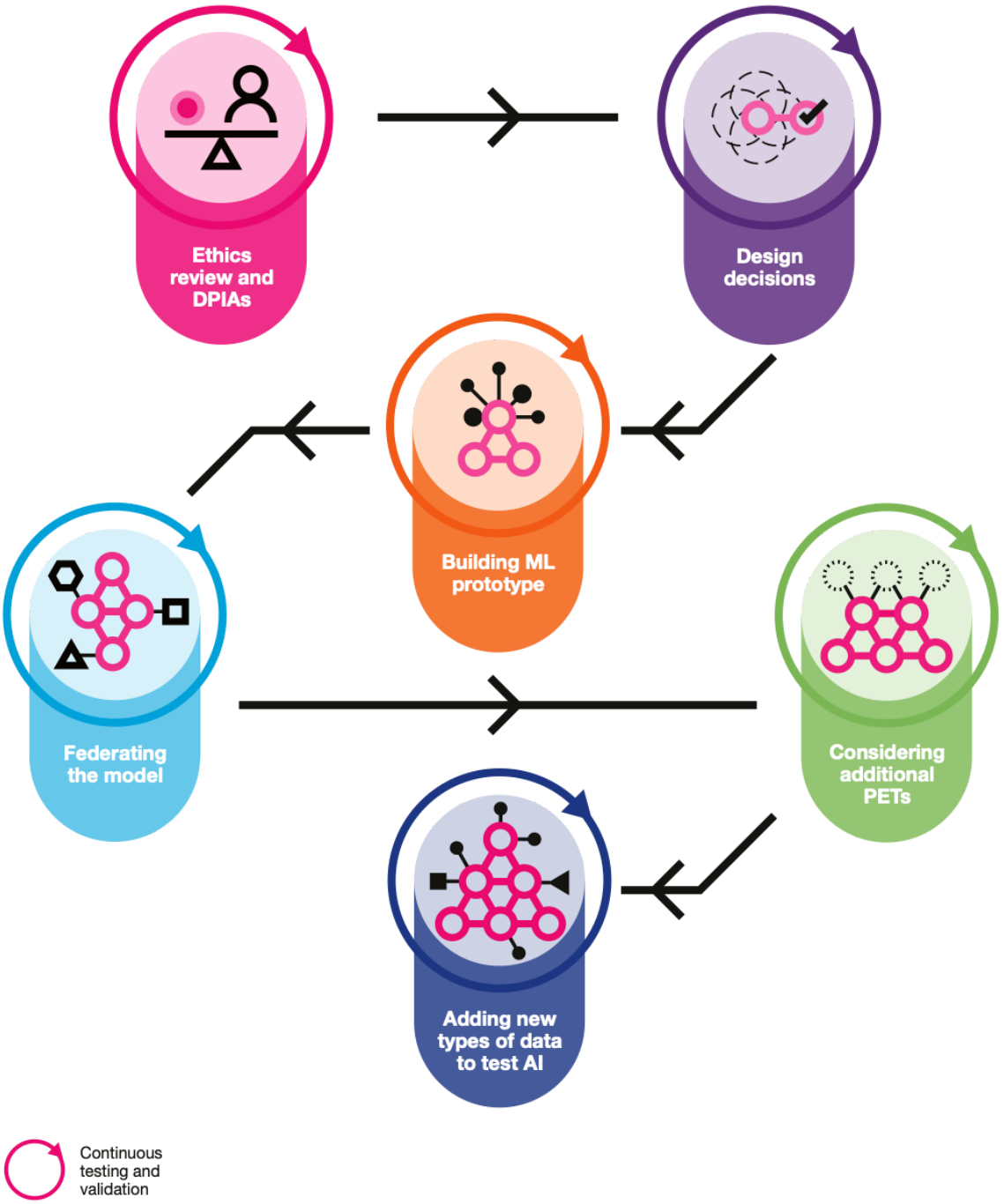
- **Number of participating organisations.** This will have several effects, such as the difficulty and time taken to coalesce around a purpose or intended outcomes and the need for processes to resolve disputes or conflicting priorities. With several organisations involved, it's important to agree on roles and responsibilities, and ways of working upfront. Individual organisations' compliance obligations, processes and policies need to be factored into the overall project plan.
- **Degree of familiarity or trust between participating organisations.** For some projects, participating organisations may have clear, common objectives. For others, it may be harder to get on the same page regarding the mutually beneficial outcomes. In these instances, an independent convenor may help broker the relationship and foster trust.
- **Homogeneity of the data and the nodes that hold it.** Federated learning governance may require deliberative processes such as agreeing on standards for data. Decisions also need to be made on the overall learning process, for example the frequency of communication between the central server and the participating nodes. Other technical processes, such as testing and validation, should be embedded throughout the deployment process to ensure that the training is going as expected.
- **Sensitivity of the data.** Data may be sensitive for reasons of privacy, commerce and/or security. Depending on the nature of the data involved, additional processes, or PETs, may be required to ensure confidentiality of data. The overall data infrastructure – and who is in control of the central server's operations – is another consideration. For highly sensitive data, penetration testing should be incorporated into the testing process.

Related to the overall project governance, it's equally important to have project management that is robust enough for the initiative to achieve its full potential. It's widely known that most IT projects fail or get severely challenged, often due to a lack of sufficient planning and inadequate management. The four complexity dimensions above would all impact the project management requirements, which may involve: gathering business requirements, managing complex and diverse sets of stakeholders, facilitating ethics reviews and DPIAs, and establishing agile software development processes. Further, when the federated learning model needs to be scaled and moved out of research and development into the wider organisation, training and change management becomes imperative to ensuring that the capability is accessible to non-expert users. A full discussion of what makes good IT project management in any given use case is beyond the scope of this report.[33]

[33] For a start, you may consider recent work on the development and deployment of general machine learning systems by Lavin et al. (2022), 'Technology readiness levels for machine learning systems'

Deployment workflow

Many of our interviewees suggested that practical deployment guidance would be helpful for organisations exploring federated learning.[34] The diagram below outlines common deployment steps and considerations at each stage. The exact steps required might differ, however, depending on the application:



The deployment steps included above are illustrative and may not occur sequentially. A federated learning pilot project will often require ongoing testing and validation to establish the feasibility of the overall design and approach, as well as to manage risks, eg around data protection. This may mean repeating some of the steps if necessary.

Figure 4: *Illustrative deployment workflow*

[34] A great overview of more technical guidance was provided by Wang et al. (2021), 'A Field Guide to Federated Optimization'

Note that some of the deployment steps will be common with any other ML implementation. For example, one must decide (either internally or with partner organisations) on the type of model, training process, parameters, and deployment plan.

The deployment steps should include:

i) Ethics review and DPIA(s):

- Consider the impact of the federated learning system on the fundamental rights and freedoms of affected groups and individuals, arising both from data collection and processing, as well as deploying and using the system in practice.[35]
- Consider, from an early stage, the legal, contractual and compliance obligations. We heard from some interviews that projects often ran into challenges in these areas as a result of the novelty of the technology.

ii) Design decisions:

- Outline and document key requirements (this may involve different client-side hardware requirements) and design decisions including the datasets, choice of parameters, choice of learning process, etc.

iii) Building ML prototype:

- Start by using public data in a centralised ML model that does not use federated learning. The objective of the prototype is to develop a better understanding of the specific project requirements.
- While using non-sensitive data, you may gain limited insights into the actual model performance and computational costs, as well as get a sense of whether ML/federated learning is suitable for the use case.
- You might, at this stage, decide not to proceed further based on the evidence generated through the process of building the prototype.

iv) Federating the model:

- Federate the model (on the same machine) and compare the model performance against the centralised model ('federated learning simulation').
- If the performance decreases substantially, you should try to identify the root cause (eg the data partitioning choice or the optimisation algorithm); usually, the use of federated learning should not cause a drop in performance. This is also the time to identify the best-suited framework for your project – or to write your own, in case the complexity of your project is limited or you are looking to stop at the prototype stage.
- Some relevant and commonly used frameworks are Flower, TensorFlow Federated, OpenMined, Scaleout and HPE Swarm Learning.
- At the same time, the rollout of simpler federation tasks to the decentralised data network (ie before training on the actual, and potentially highly complex, task at hand) helps both to validate the data assumptions and to test the implementation.

v) Considering additional PETs:

- Consider combinations with other PETs to ensure the confidentiality of data (including secure aggregation and differential privacy), and re-evaluate the

[35] Some resources that may be of use at this stage include the ODI's Data Ethics Canvas and Assessing risk when sharing data: a guide

model performance.
- Ask if additional PETs help meet the project requirements *and* whether they result in unnecessary complexity.[36]

vi) Adding new types of data:

- Start adding new types of data that are more representative of the data that will be used in production. Each new data type may come with additional challenges in terms of data access, or data protection implications.

In addition to the core steps outlined above, ongoing testing and validation should take place throughout the lifecycle of the federated learning project being deployed:

- Conduct threat modelling and explore potential attacks: federated learning has its own limitations in terms of the kind of security and privacy guarantees it provides, as opposed to traditional ML. Therefore, you should assess the exact privacy and security threats to make informed designs about using federated learning and subsequent designs. As part of this, explore potential attacks on the federated learning system before putting it into production.
- Verify the model, including preventing bias and anticipating poisoning attacks and other potential attacks.[37]
- Test how much incremental value is added throughout the training process and whether the training is going in the right direction.
- Once the federated learning system is moved into production, monitor the deployment and be prepared to handle unexpected situations (eg if a node turns out to be a malicious participant).

Lastly, we would like to reiterate that our workflow represents an illustrative scenario and that federated learning technologies are still undergoing rapid development. The exact steps that will be necessary for practical deployment might differ significantly from those provided in this section. Rather than being a definite guide, it represents a moving target.

[36] Refer to Table 3 on PETs in this report. Some helpful guidance on this can be found in the ACM article 'Federated Learning and Privacy'
[37] See the section on Security for more details.

Conclusions

On the state of federated learning:

1. **Federated learning is generating a lot of interest and attention**. This has developed due to greater focus on, and scrutiny of, data privacy – particularly when dealing with sensitive data.

2. **The value proposition for federated learning is two-fold**. Firstly, it could help organisations access more, or a greater variety of, data, which can in turn make ML models more accurate, generalisable or less biased. Secondly, federated learning helps to overcome the need to directly share data, which can be helpful when there are low levels of trust between actors or organisations.

3. **While other communities are beginning to explore federated learning, most of the focus currently remains within the academic community**. As a result, we are seeing many prototype applications being proposed and lively debates around associated dimensions of the technology. Yet, many questions around governance, compliance and ethics remain unaddressed.

4. **At the time of writing, we are at a critical juncture for federated learning and PETs more broadly**. Federated learning presents a means to generate more value from sensitive data than is possible at present. However, there is a risk that federated learning and similar technologies are deployed primarily by large companies as a means of personalising digital services and increasing profits, ahead of applications across multiple organisations to address wider societal challenges.

5. **There are two main architectural types of federated learning being used or trialled: cross-device and cross-silo**. The former tends to learn from a large number of somewhat homogenous devices, whereas cross-device federated learning is suited to deriving insights from fewer – and more heterogeneous – devices, often across organisations.

6. **Cross-device applications of federated learning are the more established of the two, but the potential of cross-silo applications is gaining traction**. Relatively few examples of fully deployed cross-silo federated learning exist at present, but a number are now being trialled.

7. **Emerging examples of cross-silo use cases of federated learning are heavily focused on leveraging the value of health data**. The use of federated learning appears to be a natural fit to overcome the challenges faced by healthcare providers. However, other uses lack as much evidential depth in the form of examples.

8. **Federated learning has reached a level of maturity at which it can often achieve similar levels of performance to traditional ML**. While maturing as a technology, the know-how and confidence to apply federated learning still lags behind, in part due to the relatively few tried and tested examples.

9. **Federated analytics is an adjacent technology that is generating significant interest, however practical examples and business cases for the technology remain in their infancy.** Our research suggests that federated analytics is now receiving similar, if not greater, attention than federated

learning. The two are related: federated analytics can help provide evidence of the potential benefits of federated technologies for an organisation, as the technology is intended to conduct simple data analyses, before they may choose to attempt the more complex training of ML models.

10. **Other existing methods for conducting ML or data science might be better suited for different audiences or circumstances**. This observation is especially true for federated analytics: distributed/decentralised analytics has been in existence for a while.

11. **Important ethical challenges remain around the use of federated learning**. For example, the phenomenon of bias is a huge challenge for ML generally. In the case of federated learning, previous work on bias mitigation is not only much less developed than for general ML, but also suggests that it is much harder to accomplish, due to lack of access to the underlying data to assess and address its biases.

Guidance and considerations around the deployment of federated learning:

12. **At present, the greatest apparent benefit from using federated learning is increasing the scalability of the ML process – not privacy preservation**. As evidenced by applications such as Google Gboard, training can be conducted over millions of devices while reducing the need to share sensitive data in plaintext. Opportunities exist to apply this approach to sensors, smartwatches and other IoT devices, which could help make models more accurate and generalisable.

13. **There remains uncertainty around the obligations under various types of law, including copyright (eg who owns what parts of a trained model); the work on these legal questions is still mostly academic and has not been extensively or consistently verified by other parties, eg in courts**. According to our research though, the obligations under GDPR, as they pertain to data controllers and processors, would usually apply in federated learning settings, as long as personal data is processed. While federated learning is still undergoing heavy development and changes, it might be best to assume that the traditional compliance requirements for development and deployment of federated learning apply regarding contractual agreements between participating parties.
14. **Federated learning, when used in isolation, is not a cure-all solution for data privacy concerns**. Federated learning is not inherently privacy-preserving, as the opportunity remains for malicious actors to infer data characteristics from model updates if additional privacy technologies are not used in conjunction. It is therefore important for potential deployers to consider what additional measures they need to take in order to ensure that data is sufficiently protected, such as the use of additional privacy technologies.

15. **The use of federated learning can pose unexpected and unquantifiable privacy and compliance risks, which can prove challenging to the use of federated learning in a collaborative fashion between different actors (eg in a cross-silo fashion)**. These challenges are similar to traditional ML, for which reliable measures of risk tend to be absent. The risks can be – to some extent – mitigated by combining federated learning with other PETs – which in turn, can also create a false sense of security. That said, actors may differ in their estimation of what can be considered as acceptable levels of risk – particularly when dealing with high degrees of uncertainty. This can inhibit the

development and deployment of federated learning while consensus is reached on what is an acceptable level of risk for all parties.

16. **Increasing awareness and knowledge of federated learning, via this type of guidance, is one way we hope to drive its use for public, educational and charitable aims**. This ambition will require substantial effort by policymakers, regulators, industry practitioners, academics and not-for-profits. There is much more to be done to stimulate experimentation and deployment of federated learning and other new PETs towards social impact.

17. **More work is needed to imagine governance models that could provide checks and balances on the power of the handful of technology companies that currently dominate the federated learning infrastructure landscape**. At the moment, we see a dominance by a few companies in cross-device federated learning. Since federated learning is still developing and changing quickly, this might still change. Meanwhile, dominant governance models have not yet emerged in the cross-silo federated learning space. Further research is required to imagine what types of governance models would be most suitable and how to implement them in practice.
18. **While federated learning is far from perfect, it strikes an important middle ground between common models of decentralised and centralised data processing**. As such, federated learning emerges, like the political decision-making of federations of nation states, as a suitable means to manage and distribute power: while a centralised model brings plenty of potential for abuse (as in authoritarian states or traditional ML), fully decentralised models are often too inefficient. In other words, federated learning can serve as an important means in democratising data use – if responsibly deployed.

Endnotes

About the Open Data Institute

The ODI is a non-profit with a mission **to work with companies and governments to build an open, trustworthy data ecosystem**. We work with a range of organisations, governments, public bodies and civil society to create a world where data works for everyone.

Our work includes applied research, consultancy services, training, and providing free reports, tools and webinars.

Authors

This report was written by Anastasia Shteyn, Konrad Kollnig and Calum Inverarity. The authors are grateful for the contributions made to the creation of this report by the wider internal team at the ODI, including Jack Hardinges, Jared Keller and Yusuff Adigun. Thanks are also due to our former colleague, Aditya Singh, who contributed to the early stages of this work.

Methodology

This report is intended to be used as practical guidance for organisations to use when exploring or experimenting with federated learning.

Our research began with a desk research phase, in which we carried out a literature review on federated learning to understand its benefits, drawbacks, maturity and prospects.

We then undertook 18 interviews with technical, legal and governance experts familiar with federated learning and related PETs. These interviews also included discussions with individuals from organisations experienced in the development and deployment of federated learning, who were able to provide much of the hands-on experience that informed the practical guidance section of this report. Interviewees were identified through the initial desk research phase, as well as through the ODI's network, before utilising 'snowball sampling' to follow up on specific areas of inquiry raised during the interviews.

We drafted an interim research findings note, which was circulated for comment with interview participants and other stakeholders, and formed the basis of a 'red teaming exercise', in which the major findings were subject to external scrutiny in a workshop setting. We then carried out further research to address areas raised during the red teaming session to inform the drafting of the final report.

Project participants

Nitin Agrawal
Reuben Binns, University of Oxford
Sandra Carrasco Limeros, AI Sweden
Jon Crowcroft, The Alan Turing Institute
Siddhartha Datta, University of Oxford
Emiliano De Cristofaro, University College London
Hamed Haddadi
Alex Ingerman
Pearse Keane, Moorfields Eye Hospital
Nicholas D. Lane, University of Cambridge
Matt Prewitt, RadicalXChange Foundation
Stephanie Rossello, KU Leuven
Lisa Sjöblom, the AI Competence Center, Sahlgrenska University Hospital, Region Västra Götaland
Juulia Suvilehto, the AI Competence Center, Sahlgrenska University Hospital, Region Västra Götaland
Naaman Tammuz, Bitfount
Blaise Thomson, Bitfount
Andrew Trask, OpenMined and University of Oxford

Acknowledgements

We would like to thank the Rockefeller Foundation for providing the opportunity to conduct this research, and all of our project participants, who were generous with their time and expertise. Thanks are also due to the wider ODI team who participated in discussions around federated learning in a variety of forums.