



Technology Safeguards for the Re-Use of Confidential Data

Application of Federated Learning for National Forest Inventories

HIGHLIGHTS

To fully exploit the economical and societal value of confidential data and avoid incurring opportunity costs, strategies to unlock their re-use are increasingly needed.

Re-using confidential data requires safeguards to protect the data and establish trust. In this document, the focus is on technological safeguards that can be deployed in federated settings. Such safeguards are illustrated for the case of National Forest Inventories and can be generalised to other cases and data sets.

Data Visiting is a technological paradigm in which holders can retain full control of their confidential data. Algorithms are shared with data holders to be run on data, and only results (not data) are shared.

Advantages and disadvantages of Data Visiting are discussed, as well as potential risks and available countermeasures. Additional safeguards can be used jointly with Data Visiting strategies to further protect data and data holders, thus reducing inherent risks.

RE-USE OF CONFIDENTIAL DATA: WHY IT MATTERS

In today's digital age, vast amounts of data are constantly created and collected. Some data are openly shared and carry no restrictions on their use or processing, while others are protected and subject to restrictions or inaccessible for re-use by third parties. This incurs *opportunity costs*, i.e., losses from not tapping on the potential value of data in known applications. Data to be protected can be personal or non-personal. Sensitive personal data are protected by the EU GDPR (2016/679). Sensitive non-personal data are referred to as *confidential* in this document. They are often deemed so due to commercial or intellectual property constraints, in addition to operational reasons.

Take the case of **National Forest Inventories** (NFIs). NFIs collect and store data coming from the long-term monitoring of forests. NFI authorities use both temporary field plots (measured only once) and permanent plots (fixed locations) to monitor the health of a forest.

Figure 1 - measurements carried out in forests.



source: curated by [Valerief](#) on [Unsplash](#)

Institutions in charge of NFIs consider that the location of permanent plots should not be shared to avoid changes in the forest management, which could bias or otherwise compromise the statistics derived from the data. Hence, the locations of the permanent plots are considered *confidential statistical units* and access by third parties is limited or

blocked. However, restricting access to plot-level NFI data incurs *opportunity costs*. NFIs' ground-based inventories can be used to calibrate and validate data products and models generated through remote sensing, such those of Copernicus, the Earth Observation (EO) component of the European Union's space programme. The joint use of NFI and EO data could significantly improve forests health information and help adapt forest management practices to the climate crisis.

Forests are not the only example. Take, for instance, the case of **health data**. While it is important to share clinical records with medical research centres and hospitals to advance research, at the same time it is crucial to protect patients' confidentiality and avoid identification risks.

An additional example comes from **agriculture**. Sharing farm data enables the development of increasingly accurate services for farmers to support sustainable production, but it is paramount to avoid disclosing sensitive business information.

EU POLICY CONTEXT

The 2024-2029 political guidelines of President Ursula von der Leyen highlight that "*Europe needs a data revolution*", emphasising the importance of accessing data in a trustworthy and secure manner to support innovation. The policy context for re-using protected data -including those with confidentiality constraints- has become more favourable in the last few years.

The Data Governance Act (DGA) (2022/868) acknowledges the potential value lying in the "re-use of certain categories of protected data held by public sector bodies". Specifically, Chapter II of the DGA, in Articles 3-9, outlines the categories of protected data, which includes confidential data, the conditions for re-use, the safeguards that need to be in place, and the procedure for requesting access.

The proposed Monitoring Framework for Resilient European Forests (**Forest Monitoring Law 2023/0413**) aims to take stock of the provisions of the DGA on confidential data owned by the public sector. Article 7 focuses on the establishment and use of (technological and operational) safeguards that will allow reusing data from NFIs permanent plots without compromising their confidentiality. In addition, Article 9 states that the Commission is empowered to establish *confidentiality-preserving*

safeguards for the inclusion of monitoring sites data in the Forest Information System.

Confidentiality Preserving Safeguards are also instrumental to support data sharing initiatives, such as the GreenData4All initiative, which aim to facilitate the sharing of environment-related data in the context of the European Green Deal. Furthermore, the design and deployment of proper safeguards is crucial for the success of the Green Deal Data Space (GDDS) and all the Common European Data Spaces (CEDS). Secure and confidentiality-preserving instruments serve as key enablers to empower data holders, thus fostering data availability and facilitating exchanges.

USING PROTECTED DATA WITHOUT COMPROMISING CONFIDENTIALITY

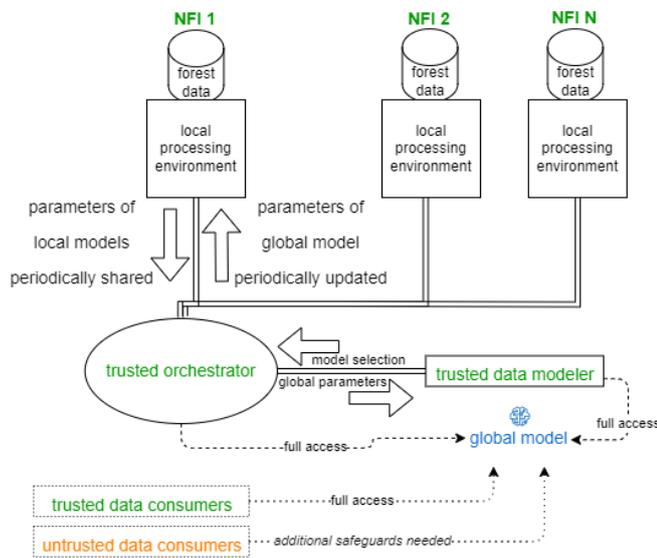
This policy brief proposes solutions that can **extract value from the data without compromising their confidentiality**. Although there is no single technical or organisational approach that can be applied to every case, mature tools are available and can be used in different contexts.

Privacy-Enhancing Techniques (PETs) can be considered key enablers in this regard. Their use requires accepting a trade-off, because protecting data has a cost in terms of lowered accuracy of the results, increased computational costs, and so on.

From Data Sharing to Data Visiting

Using protected data without compromising confidentiality requires the use of PETs. Among such solutions, the Data Visiting concept represent a paradigm shift: instead of moving the data from the holder to the user to be algorithmically analysed, it moves the algorithm to the data. Such paradigm is referred to as **Data Visiting** as opposed to **Data Sharing**. Data visiting thus foresees that an algorithm is shared, manually or automatically, with data holders, for it to run in a secure local processing environment with access to confidential data. The results (not the data) are then shared with the data users. In this way, the data holder never loses control of data.

Figure 2 - NFIs jointly train an AI-based model that can infer forest indicators. The training is orchestrated by a trusted entity. Results, which are model parameters (not NFI data), are then shared with data consumers, such as the data modeler.



source: own production

Figure 2 illustrates a data visiting scenario in which NFI data are used to train an AI model. NFI authorities stay in control of their data, which are not shared with others. The involved parties are:

- Data Providers:** in this case, trusted NFI authorities, which do not share confidential data but allow algorithms to be run locally on them;
- Orchestrator:** assumed as trusted, in charge of the infrastructural (hardware and software) components used to manage the process in collaboration with the NFIs, which use local processing environments;
- Data modeler:** in charge of selecting the model to be trained and defining the indicator(s) of interest to be derived; could be trusted with full access to the global model (as in fig. 2), or untrusted and thus requiring the deployment of additional safeguards;
- Data consumers:** interested in the final results, have no active role in the process; could be trusted with full access to the global model, or untrusted thus requiring the deployment of additional safeguards.

A trusted entity, such as the Joint Research Centre (JRC) of the European Commission or the European National Forest Inventory Network (ENFIN), can act as orchestrator and/or as data modeler.

The Data Visiting strategy in Fig. 2 is based on Federated Learning (FL), a privacy-preserving

technique. FL has been selected because it is a mature solution, available open source, which can be used in conjunction with additional safeguards for further protection.

What is Federated Learning

The FL technique has been proposed by Google in 2017 to train a language model, used by the Google Keyboard (GBoard), orchestrating users' mobile devices around the world; since then, several different applications have been developed, such as AI models for cancer treatment without sharing patients' confidential data.

In the NFI case, the idea is that several data holders, e.g. NFI administrators (acting as *FL clients*), jointly train an AI model to be used by one or more data consumers and by NFIs themselves. The process is iterative, which means that several rounds of exchanges are carried out among the clients and the orchestrator. Once completed, the FL clients and the data modeler have the global model. Data consumers may request a copy of it.

Strengths and weaknesses of FL

Data sovereignty (strength)

The use of FL implies that data are not shared, thus data holders remain in full control of their assets. It strongly limits the possibility that third parties access confidential data without explicit authorisation by data holders.

Data value creation (strength)

Even if confidential data are not shared, FL offers the possibility to use them and create value. In fact, the global model generated by a FL process can be used to obtain results as if data were publicly available. The generated model can be publicly shared on a case-by-case basis or not at all to further protect it and the underlying data.

Accuracy of the results (weakness)

Compared with results obtained in data sharing scenarios, FL may provide less accurate results. As a general rule, **the higher the desired confidentiality, the lower the attainable accuracy** (or utility): it is an unavoidable trade-off between confidentiality preservation and accuracy.

Threat of data reconstruction (weakness)

Even though only model parameters (and not data) are shared in a FL process, there is a residual risk that data can be partially or fully reconstructed. Attack techniques, known as *model inversion*, have

been described in the scientific literature, such as Deep Leakage from Gradients (DLG) and variants, or Inverting Gradients (IG) attacks. The reconstruction could be carried out by a malicious or an honest-but-curious attacker, who has access to the parameters exchanged among clients and orchestrator. To mitigate such risk, we suggest the use of **Differential Privacy** later in this document.

Threat of data/model poisoning (weakness)

Attackers may want to poison the global model instead of reconstructing the input data. Model poisoning occurs when its parameters are intentionally modified and compromised. Alternatively, data can be poisoned, which means that one or more datasets are compromised to introduce misleading or incorrect information in the global model. Because of poisoning, results generated from the global model are different from what is expected.

RISK ANALYSIS FOR CONFIDENTIALITY BREACH

Federated Learning and the underpinning data visiting principle offer strong and reliable safeguards. However, risks for confidentiality remain and must be carefully considered and mitigated with **additional technological and operational safeguards**. Additional safeguards should be proportional to the sensitivity of the data. A residual risk, after risk control, remains.

To guide the discussion on confidentiality risks, we formulate the following questions and related answers:

Q1: Can data be reconstructed and/or leaked, even partially, when using Federated Learning?

In the NFI case (Figure 2), both the orchestrator and the data holders are public entities, thus are likely to be trusted actors. Trust is not only based on reputation or legal status but also on the use of **robust and reliable infrastructures** to run FL, as well as of other **operational safeguards** (e.g., restricted access to facilities, authentication, biometrics, and encryption). Actors (curious modelers, data consumers or malicious third parties) might attempt to reconstruct confidential data using the model parameters exchanged among data holders and orchestrator.

Q2: Which safeguards and attack mitigation techniques can be used against threats?

External **auditors** can be employed to test the robustness of the operational and technological safeguards that the orchestrator and the data holders put in place in order to instil trust. Checks will concern not only software but also the overall production environment. The NIS2 Directive (2022/2554) can be considered as a reference, requiring the set-up of an ICT risk management framework, the use of reliable and resilient ICT systems, and the prevention of breaches of confidentiality. ISO/IEC 27001 (Information Security Management) and ISO/IEC 27002 (Information Security Controls) are well-known international standards offering a systematic approach for managing information security risks and best practices for selecting and implementing security controls.

NFIs, orchestrator and data modeler should be **authenticated**. **Data encryption** should be used to mask data exchanges among data holders and orchestrator. Additional safeguards, such as **Differential Privacy**, can be used in the presence of untrusted participants.

Finally, **testing** (e.g., hackathons) can be used to test the feasibility of data reconstruction. Results can inform all parties on ways to detect and identify risks, and suggest strategies to have appropriate mitigation actions in place.

Q3: What is the cost of Federated Learning and additional confidentiality safeguards?

Operating costs are associated with the use of Federated Learning and additional safeguards. For FL to be used, data holders must preliminarily **harmonise data** according to a common model. Furthermore, **software deployment** is expected at the orchestrator and at each data holder, which incurs costs. **Software auditing** should be added to potential cost sources, as well as the need of a **secure processing environment** at each site. Looking at results, the achievable **accuracy** is lower if compared to the case in which data are shared. The loss in accuracy depends on several factors, and the deployment of additional safeguards (e.g., Differential Privacy) may further lower accuracy.

In the following sections, we provide a more complete overview of Federated Learning and additional safeguards that can be used in conjunction with it. The NFIs case has been used so far as example scenario; in the following, we use instead a generic scenario to show how the proposed strategy can be used in different cases.

First, we introduce the concept of **Trustworthy Federated Learning**, describing desirable dimensions to set up a reliable system with accountability in mind. Then, **Differential Privacy** is presented as potential solution to be used in conjunction with Federated Learning to further protect data in the presence of untrusted parties. To conclude, and in order to provide a complete overview, also **adversary models** are presented along with practical tips to **strengthen** a real-world **FL deployment**.

TRUSTWORTHY FEDERATED LEARNING

In 2019, the Independent High-Level Expert Group on AI set up by the European Commission, presented the *Ethics Guidelines for Trustworthy AI*. Recently, Sanchez et al. extended the guidelines to propose Trustworthy Federated Learning (TFL) because they consider **trustworthiness** as a critical feature to address confidentiality concerns when using Federated Learning. The authors identify key pillars and metrics to evaluate and quantify the trustworthiness of FL.

Confidentiality

A prominent factor that has been fuelling the use of FL. The objective is to make information units increasingly indistinguishable so that re-identification is increasingly hard to achieve. Being aware that risks may come from involved parties or external actors, it is paramount to prevent threats and put in place mitigation measures. To achieve that, PETs can be used.

Robustness

In the context of FL, robustness translates into resilience to attacks and into reliability of the used infrastructure. The former means that mechanisms are in place to detect malicious actions and mitigate or negate them. The latter is related to the reputation of participating clients (data holders) and quality of the used data; the accuracy of the results (performance of the model); and the resiliency of the infrastructure used to run FL in production environments.

Fairness

Data used to train an AI model must be representative of the entire phenomenon under analysis. Discrimination and differentiated treatments should not be possible. To achieve that, all data holders are supposed to contribute, at some stage, to make sure that their data are represented, and all cases are considered. Even though underrepresentation and imbalance cannot be fully avoided in every case, the more uniform the accuracy of the global model on local data is, the fairer the global model will be.

Explainability and Interpretability

FL is an AI-based process, and transparency is often a key concern. It should be possible for humans to describe how the model works and to interpret results explaining -at least to some extent- how the model proceeded to calculation. The use of models *interpretable by design* should be preferred, also in light of the provisions on risks set in the EU AI Act (2024/1689).

Accountability

The use of FL requires multiple parties to collaborate to set up a system and train a global model. Each party is responsible for documenting actions and claims for transparency reasons. External auditors (service providers acting as intermediaries) can verify what each party has declared to foster trust and ensure accountability. Audit can also cover software components to make sure standards are followed and legal agreements are met.

Federation

Federated Learning is a collective task among several parties and therefore the management and governance of the participants' federation is utterly important. The use of registries to keep track of participating data holders, hardware and software settings in use, chosen algorithms and aggregation strategies is important for management purposes and to assess performance and robustness.

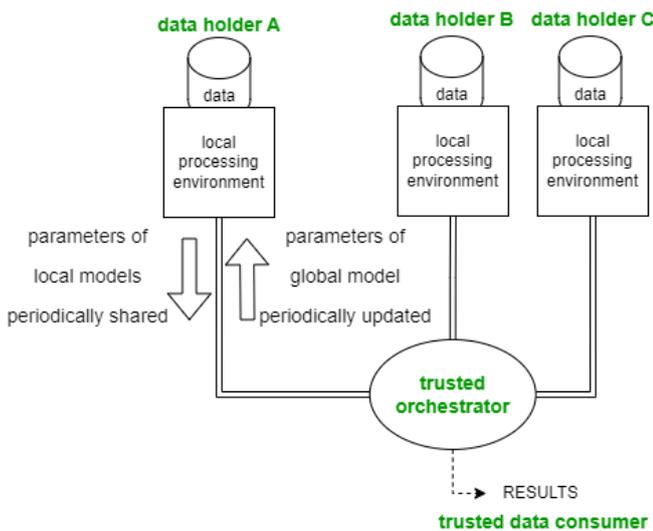
WHAT IS DIFFERENTIAL PRIVACY

Differential Privacy (DP) is a mechanism to protect confidential data by perturbing them with noise. Often used to protect data, such as census results, when they are publicly released. The approach is based on the use of carefully calibrated noise to minimise the impact on utility while lowering the risk of disclosure of sensitive information.

DP can be used jointly with FL to further protect sensitive data from the threat of reconstruction. In the figures below, the use of FL and DP is shown in three different scenarios, under the assumption that the data modeler acts as orchestrator too:

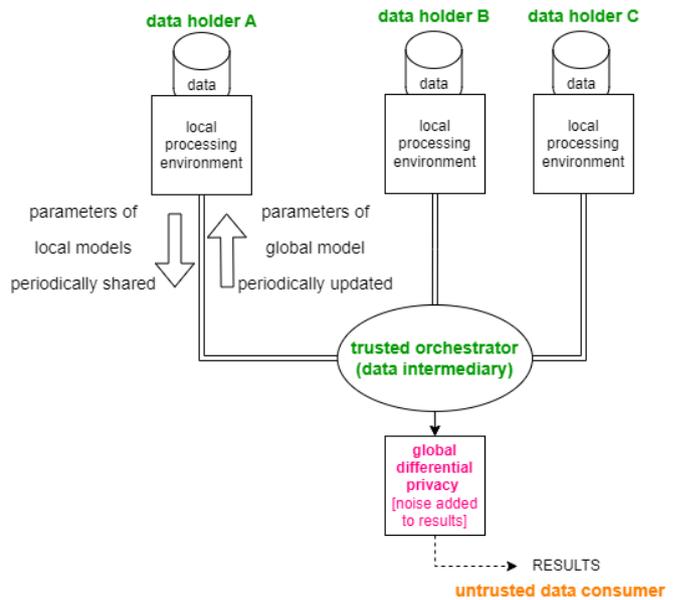
1. both orchestrator and data consumer(s) are trusted parties, thus the use of **DP may not be necessary** (see Fig. 3);
2. the orchestrator is trusted and the consumer(s) untrusted, thus **DP can be used at global level** to add noise to the resulting model parameters (and not to data) (see Fig. 4);
3. both orchestrator and consumer(s) are untrusted, thus **DP is used at local level** (by each data holder) to add noise to the data before use (see Fig. 5).

Figure 3 – Use of FL with orchestrator and consumer(s) as trusted parties (DP is not used).



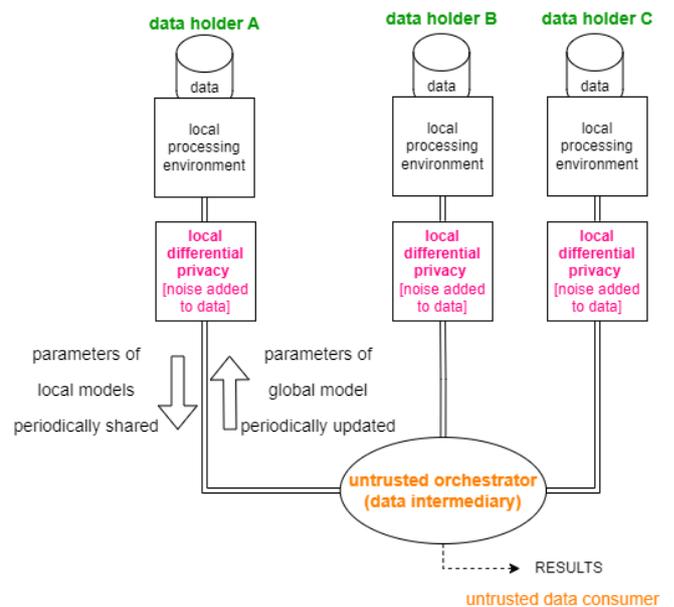
source: own production

Figure 4 – Use of FL with trusted orchestrator and untrusted consumer(s) as trusted parties: DP is used on results and protect confidentiality.



source: own production

Figure 5– Use of FL with untrusted orchestrator and untrusted consumer(s) as trusted parties: DP is used on both data and results to protect confidentiality.



source: own production

Strengths and weaknesses of DP

Confidentiality protection (strength)

DP adds a layer of protection on data or on results from the threat of reconstruction. Adding noise makes the data units increasingly indistinguishable, and the more noise is added the more indistinguishable the data units become.

Accuracy of the results (weakness)

However, adding noise entails lowering the accuracy of the results. Again, the higher the desired confidentiality, the lower the attainable accuracy: each additional measure put in place has a cost in terms of trade-off between confidentiality preservation and accuracy.

ADVERSARY MODELS (ATTACKERS)

Actors trying to breach confidentiality or poisoning data and results are referred to as *adversaries* or *attackers* in the scientific literature. Their posture can be described as passive or active.

It is worth stressing that, while there is no absolute guarantee that confidentiality will not be breached, the use of specific countermeasures, such as Differential Privacy, can lower such risks and related negative impacts. Defence from poisoning attacks requires the use of more elaborated approaches (model analysis, byzantine robust aggregation, and verification-based methods) which may prove less straightforward to implement. Results generated by poisoned data and/or models are different from what is expected; it is worth remarking that breach of confidentiality is not the key objective of poisoning.

Passive attackers

Passive attackers do not interact with other participants during the training phase. Their aim is to reconstruct data by listening to shared model parameters, which may potentially leak information under some conditions.

Active attackers

Active attackers interact with participants in the training phase. Their aim is to poison the model or the data, or to reconstruct data by altering the value of the parameters exchanged among clients.

Attack surface

The key strategy to minimise the likelihood of attacks is to limit the attack surface, i.e., entry points that attackers can exploit. For instance, the aggregation algorithm used by the orchestrator should be able to deal with data that may lead to re-identification, such as outliers.

HOW TO STRENGTHEN A FL SETUP

Client authentication

A simple yet effective strategy to limit risks is **authentication**. Only trusted parties should join FL after successful authentication. The most commonly used FL frameworks, including open source solutions, offer such capabilities.

Data Encryption

An additional means of protection is data **encryption**. All exchanges among participants should be encrypted, thus making attacks increasingly difficult to be carried out. Encryption and decryption operations to be carried out at each end at each round do increase computational load however.

Homomorphic Encryption

An encryption technique attracting interest is (fully) homomorphic encryption. Its main advantage is that it removes the need to encrypt and decrypt information; in fact, operations are not carried out on data in the clear, but **everything stays encrypted**, including results after operations. However, the significant computational load, which could be a severe performance bottleneck, and the limited adoption of the technique at present may not meet the needs of practical applications.

CONCLUSIONS

In this document, Federated Learning has been presented as a Data-Visiting-based solution to be used in real scenarios. Advantages and disadvantages have been discussed, as well as strengths and weaknesses. In our view, expected benefits outweigh the costs, and security concerns can be managed or minimised if robust and reliable procedures are in place.

REFERENCES

Päivinen, R., Astrup, R., Birdsey, R.A. et al. "Ensure forest-data integrity for climate change studies", *Nature Climate Change* 13, 495–496 (2023).

<https://doi.org/10.1038/s41558-023-01683-8>

Rieke, Nicola, Jonny Hancox, Wenqi Li, Fausto Milletari, Holger R. Roth, Shadi Albarqouni, Spyridon Bakas et al. "The future of digital health with federated learning", *NPJ digital medicine* 3, no. 1 (2020): 1-7.

<https://doi.org/10.1038/s41746-020-00323-1>

Auñón García, J.M., Hurtado Ramírez, et al., "Evaluation and utilisation of privacy enhancing technologies - A data spaces perspective", *DATA IN BRIEF*, 55, 2024.

<https://doi.org/10.1016/j.dib.2024.110560>

Konečný, Jakub, H. Brendan McMahan, Daniel Ramage, and Peter Richtárik. "Federated optimization: Distributed machine learning for on-device intelligence", arXiv preprint arXiv:1610.02527 (2016).

<https://arxiv.org/abs/1610.02527>

"Federated Learning: Collaborative Machine Learning without Centralized Training Data"

<https://research.google/blog/federated-learning-collaborative-machine-learning-without-centralized-training-data/>

Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith. "Calibrating noise to sensitivity in private data analysis" In *Theory of Cryptography: Third Theory of Cryptography Conference*, New York, NY, USA, March 4-7, 2006. Proceedings 3, pp. 265-284. Springer Berlin Heidelberg, 2006.

https://doi.org/10.1007/11681878_14

Sánchez, Pedro Miguel Sánchez, et al. "FederatedTrust: A solution for trustworthy federated learning" *Future Generation Computer Systems* 152 (2024): 83-98.

<https://doi.org/10.1016/j.future.2023.10.013>

"Ethics guidelines for trustworthy AI"

<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Xia, G., Chen, J., Yu, C., & Ma, J. (2023). "Poisoning attacks in federated learning: A survey" *IEEE Access*, 11, 10708-10722.

<https://doi.org/10.1109/ACCESS.2023.3238823>

Gentry, Craig, Amit Sahai, and Brent Waters. "Homomorphic encryption from learning with errors: Conceptually-simpler, asymptotically-faster, attribute-based", 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part I. Springer Berlin Heidelberg, 2013.

https://doi.org/10.1007/978-3-642-40041-4_5

COPYRIGHT

© European Union, 2025, except: Figure 1 © [Valerief - Unsplash](https://www.unsplash.com).

How to cite: European Commission, Joint Research Centre. Bacco, M., Kanellopoulos, S., Di Leo, M., Kotsev, A., Friis-Christensen, A., *Technology Safeguards for the Re-Use of Confidential Data*, European Commission, Ispra, 2025, JRC141298.