Introducing the TRUMPET project: TRUstworthy Multi-site Privacy Enhancing Technologies

Alberto Pedrouzo-Ulloa[‡], Jan Ramon[§], Fernando Pérez-González[‡], Siyanna Lilova[‖], Patrick Duflot[**]
Zakaria Chihani[††], Nicola Gentili[¶], Paola Ulivi[¶], Mohammad Ashadul Hoque[‡‡], Twaha Mukammel[‡‡]
Zeev Pritzker[†], Augustin Lemesle[††], Jaime Loureiro-Acuña[*], Xavier Martínez[*], Gonzalo Jiménez-Balsa[*]

[*] Galician Research and Development Center in Advanced Telecommunications (GRADIANT)
Vigo, Spain, Email: {jloureiro, xmartinez, gjimenez}@gradiant.org

[†]Arteevo Technologies
Tel Aviv, Israel, Email: zeev@arteevo.com

[‡]atlanTTic Research Center, Universidade de Vigo
Vigo, Spain, Email: {apedrouzo, fperez}@gts.uvigo.es

[§]Institut National de Recherche en Informatique et Automatique (INRIA)
Lille, France, Email: jan.ramon@inria.fr

[¶] Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori (IRST) IRCCS
Meldola, Italy, Email: {nicola.gentili, paola.ulivi}@irst.emr.it

[‖]Timelex
Brussels, Belgium, Email: siyanna.lilova@timelex.eu

[**]CHU de Liège
Liège, Belgium, Email: pduflot@chuliege.be

[††] CEA-List
Palaiseau, France, Email: {zakaria.chihani, augustin.lemesle}@cea.fr

[‡‡]Technovative Solutions
Manchester, United Kingdom, Email: {ashadul, twaha}@technovativesolutions.co.uk

***Abstract*—This paper is an overview of the EU-funded project TRUMPET (https://trumpetproject.eu/), and gives an outline of its scope and main technical aspects and objectives.**

In recent years, Federated Learning has emerged as a revolutionary privacy-enhancing technology. However, further research has cast a shadow of doubt on its strength for privacy protection. The goal of TRUMPET is to research and develop novel privacy enhancement methods for Federated Learning, and to deliver a highly scalable Federated AI service platform for researchers, that will enable AI-powered studies of siloed, multi-site, cross-domain, cross-border European datasets with privacy guarantees that follow the requirements of GDPR. The generic TRUMPET platform will be piloted, demonstrated and validated in the specific use case of European cancer hospitals, allowing researchers and policymakers to extract AI-driven insights from previously inaccessible cross-border, cross-organization cancer data, while ensuring the patients' privacy.

***Index Terms*—Federated Learning, Privacy Metrics, Privacy Enhancing Technologies, General Data Protection Regulation, Clinical Cancer Research, Patient Data Privacy**

The list of affiliations follows the order indicated in the Consortium Agreement of the TRUMPET project [1], [2].

I. Introduction

Learning thrives on data. The recent renewed popularity of Machine Learning (ML) largely owes to the availability of vast training datasets. However, datasets with the required quality are often not available from a single source, and need to be assembled from subsets owned by different organizations that have a variety of access policies and that often grant no access at all to external entities, thus implementing policies that are far more stringent than required by the GDPR (General Data Protection Regulation).

In recent years, Federated Learning (FL) has emerged as a revolutionary privacy-enhancing technology [3], apparently able to address the problem of collaborative training. Popularized by Google for privacy-preserving prediction of user keystrokes on smartphones [4], FL has quickly expanded to other applications [5], [6], [7]. Among others, peer-to-peer network-based architectures have been created that eliminate FL's single point of failure by decentralizing AI (Artificial

© 2023 IEEE. Personal use of this material is permitted.
2023 IEEE International Conference on Cyber Security and Resilience (CSR) | DOI: 10.1109/CSR57506.2023.10224961

Intelligence) model aggregation [8], [9]; novel algorithms have been developed to ensure convergence of FL's global AI model and to cope with intermittent connectivity of the FL learning nodes in certain applications [10].

While FL was originally positioned as a major privacy-preserving innovation, further research has cast a shadow of doubt on the strength of privacy protection provided by FL [11]. Potential vulnerabilities and threats pointed out by researchers included a curious aggregator threat [12]; susceptibility to man-in-the-middle and insider attacks that disrupt the convergence of global and local models, or cause convergence to fake minima [13]; and, most importantly, inference attacks that aim to re-identify data subjects from FL's AI model parameter updates [14].

In view of the strong privacy protection stance of the European Union, expressed through the CyberSecurity Strategy, the CyberSecurity Act[1] and the GDPR—that stipulates penalties in the amount of 4% of global revenues of noncompliant businesses—these new findings regarding the privacy guarantees provided by FL are worrisome for business managers, and also harmful to the proliferation of FL as a technology that enables GDPR compliance.

A. The TRUMPET project: An Armored FL Framework

The TRUMPET project [1], [2] aims at building a multi-sided, privacy-enhanced, Federated Learning-based platform for the development of dedicated AI applications that will assist (in the use cases selected for the project and described in Section VI) healthcare professionals. Our platform for Armored FL (AFL) will enable AI-powered studies of siloed, multi-site, cross-domain, cross-border European datasets with privacy guarantees that follow the requirements of GDPR. The key end users are researchers/solution developers (on the demand side), data owners (on the supply side) and healthcare professionals (indirect users of TRUMPET on the application side), as shown in Figure 1.

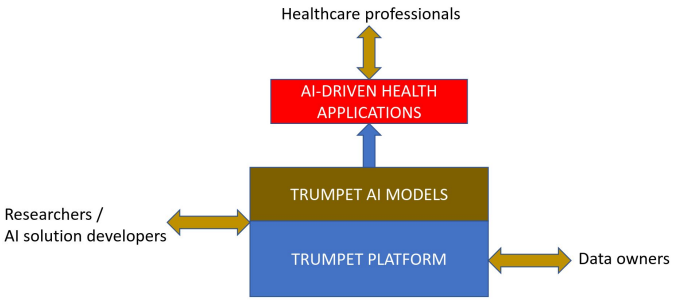


Fig. 1. TRUMPET and its direct and indirect users.

B. TRUMPET's Main Objectives

The goal of the TRUMPET project is to deliver a highly scalable Federated AI service platform for researchers (see Section II), which will enable AI-powered studies of siloed, multi-site, cross-domain, cross-border European datasets with privacy guarantees that follow the requirements of GDPR.

To this aim, the project will focus on two complementary technical aspects: (1) to research and develop novel privacy enhancement methods tailored for FL (see Section III), and (2) to research, develop and promote with EU data protection authorities a set of novel privacy metrics and a tool for the privacy evaluation of FL implementations (see Section IV).

During the project, we will identify and perform a legal study of the implications of GDPR in FL implementations, which may be applied with further restrictions by different European countries. Our final aim is to identify grey areas, where the use of PET (Privacy Enhancing Technologies) techniques and our novel privacy measurement tool could be applied to help to remove uncertainty on the GDPR compliance of Armored FL (see Section V).

The generic TRUMPET platform will be piloted, demonstrated and validated in the specific use case of European cancer hospitals (see Section VI), allowing researchers and policymakers to extract AI-driven insights from previously inaccessible cross-border, cross-organization cancer data, while ensuring the patients' privacy.

II. The TRUMPET platform

High-Level Description: Data Owners (hospitals) keep their datasets private and do not share them directly. Instead, after being given trustworthy privacy guarantees by the TRUMPET platform operator, they install a TRUMPET learning node on their premises and allow it to access the data, so it can be trained and processed on the Data Owner's premises, by a locally executed AI model.

The training of the local model can be performed either continuously or in batches, depending on the application. During training, only the parameters of the resulting model are exchanged with a central node called Federated Model Management (FMM), which is located outside the hospital's premises. The FMM aggregates the local model updates into a global one, and shares it with the TRUMPET nodes, so as to replace the local model updates and continue training if required. This FL process converges to a global model that is constructed from the private datasets of individual Data Owners, but without the data being moved across the boundaries of each Data Owner's premises.

Figure 2 illustrates the flow chain of the TRUMPET platform acting as a privacy-enforcing intermediator. Hospitals keep their data private and do not share them directly. Instead, they install a TRUMPET learning node on their premises and allow it to access the data, so it can be trained and processed on the hospital's premises, by a locally executed AI model. We refer the reader to Section VI for more details on the clinical use cases.

A. Platform Architecture

The TRUMPET platform has a distributed architecture in which: (1) patient data resides in a data owner node located inside the hospital's premises, and (2) the central TRUMPET

[1]https://digital-strategy.ec.europa.eu/en/policies/cybersecurity-act.

© 2023 IEEE. Personal use of this material is permitted.
2023 IEEE International Conference on Cyber Security and Resilience (CSR) | DOI: 10.1109/CSR57506.2023.10224961

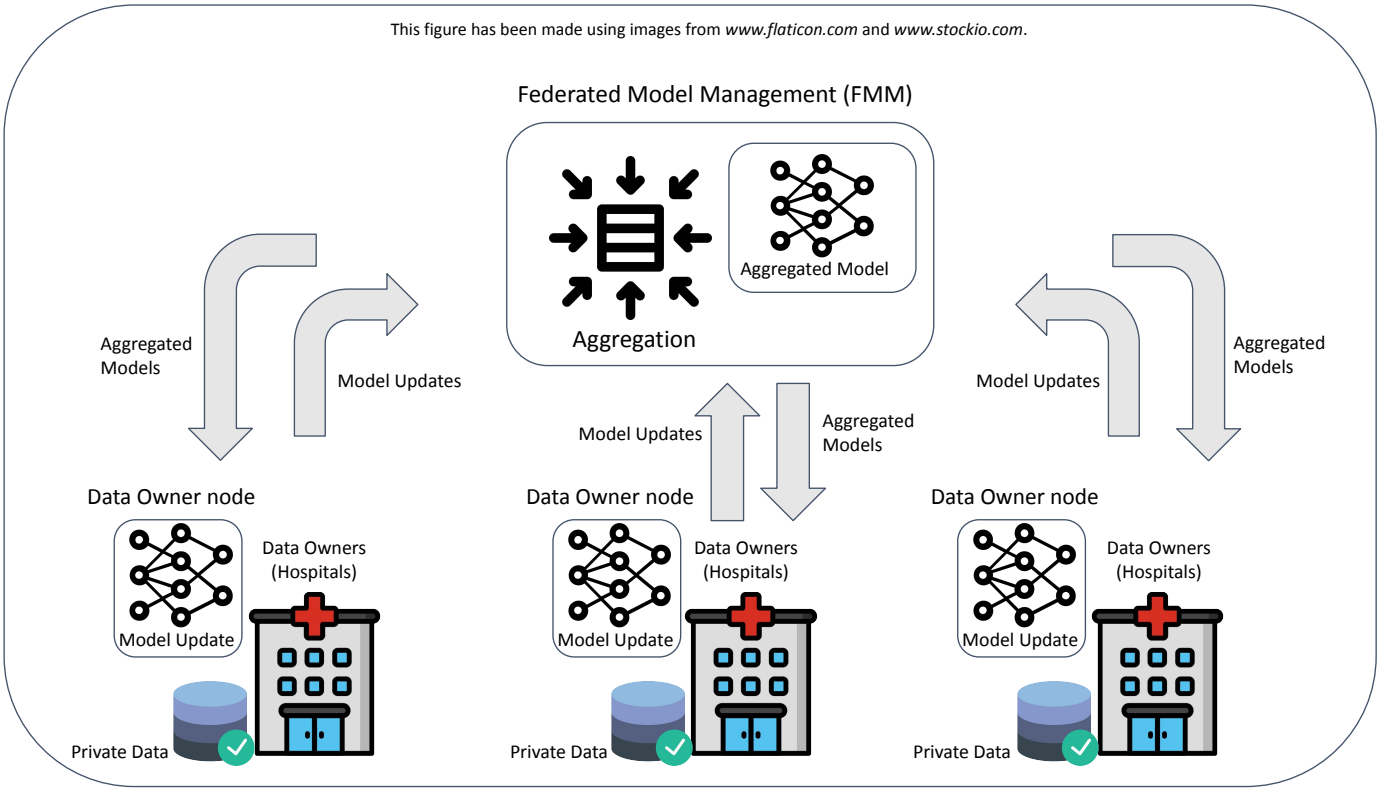


Fig. 2. High-level description of the TRUMPET flow chain.

cloud acts as the facilitator for researchers to build AI models without access to the data. It is worth noting that only descriptive metadata, summary statistics and model updates leave the data owner's premises. Figure 3 details the two main dashboard components of the TRUMPET architecture.

1) Data Owner Dashboard: The data processing and transformation (ETL; Extract, Transform, Load) service is located at the data owner node. It exposes REST APIs which are used by the hospital's IT team to upload patient datasets privately to the data owner node. Then, each data owner can describe its dataset, define the required privacy level, and also advertise it to external researchers by making use of its data owner control panel. Different versions of the ETL service are envisioned to support several ETL strategies suitable for different healthcare systems. The control panel also provides a dashboard to monitor the different advertised datasets for GDPR transparency, review the model training requests, etc.

Each data owner node will have isolated its user administration and be fully managed by the data owner's IT team. Three user groups are envisioned in the data owner node: (a) *Administrator*: performs user management, (b) *Privacy manager*: sets the privacy level of datasets, defines descriptive metadata of datasets, advertises datasets, approves data requests made by external researchers, mainly focusing on aspects related to ethical and data protection, monitors the advertised datasets, and monitors data owner node health, (c) *Observer*: monitors advertised datasets and monitors data owner node health.

2) Researcher Dashboard: The researcher dashboard will facilitate that researchers can browse advertised datasets, add/remove a dataset to/from favorites, define cohort, get descriptive statistics, train AI models, etc. For example, the model training management service will coordinate model training at data owner nodes, which is done via the agent services acting on behalf of a researcher. First, the researcher will define the code to be executed at the training agent of a data owner node. Afterwards, the code will be reviewed by the data owner and executed on the private dataset. Finally, if the code is approved by the data owner, results of the execution will be returned to the researcher.

We envision to develop a library of pre-approved code that can be executed by the training agent without data owners' intervention. This library of pre-approved code will differ from data owner to data owner, hence reflecting both the data owner's comfort level and dataset requirements. PET methods and privacy budget will be transparently applied to all code execution to prevent data leakage. Finally, the case of descriptive analytics will follow a similar behavior, but without requiring the explicit intervention of the data owner.

B. Data Collection

We will harmonize clinical datasets into a common format and inject them into the TRUMPET private repositories. With the support of technical partners, hospital partners (IRST IRCCS and CHU de Liège) will align their data with a common data model. They will also extract information from

© 2023 IEEE. Personal use of this material is permitted.
2023 IEEE International Conference on Cyber Security and Resilience (CSR) | DOI: 10.1109/CSR57506.2023.10224961

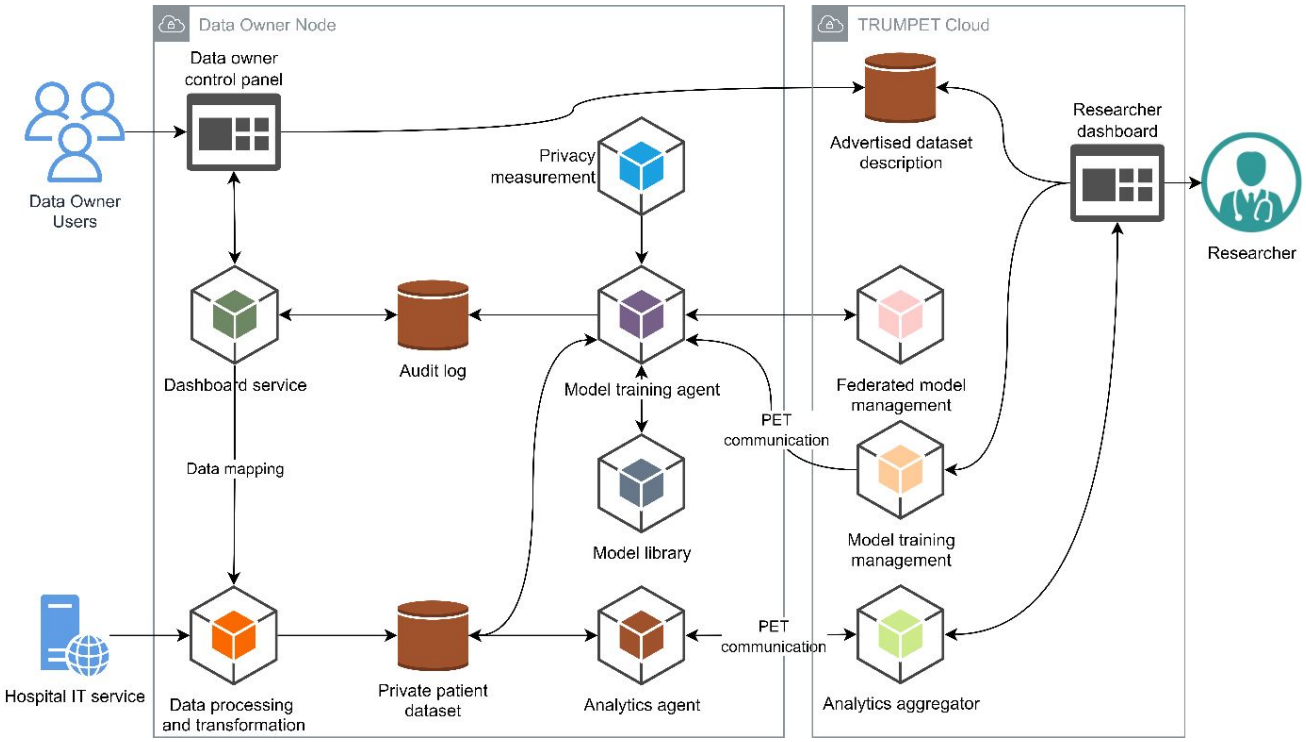


Fig. 3. High-level view of the TRUMPET platform architecture.

databases, data warehouses or other types of data sources that can be shared with the TRUMPET platform to support the three use cases. This data collection will be performed in batch mode (see Figure 4).

There will be no data continuously streamed from the clinical databases to the TRUMPET data owner node. It shall also be understood that only data for those patients matching the inclusion and exclusion criteria of the TRUMPET use cases will be extracted. Moreover, the complete patient record will not be extracted. Only a list of variables needed to support the use cases will be extracted. A detailed description of the three use cases is included in Section VI.

C. Platform Development and Validation

The TRUMPET platform will be validated in three clinical use cases which are detailed in Section VI. In contrast, some more general aspects regarding the development and validation of the TRUMPET platform are included here.

1) Two-Round Pilot Approach: The prototyping and piloting in all use cases will follow an interactive and incremental approach to ensure continuous improvement of the TRUMPET platform throughout the project. The pilot will be divided in two rounds associated to two different milestones. In the first round, an initial-basic prototype will be developed and piloted to get preliminary feedback on the platform features. These first indicators will ensure early identification of needs for further adjustments. They will also facilitate the fine-tuning of the initial prototype during the second round, by making appropriate technical improvements according to the results and potential new requirements elicited in the first round.

2) Independent Penetration Testing: With the aim of fostering trust of data owners in the TRUMPET platform, the project will (1) engage external penetration test experts to validate the privacy protection, and (2) provide a PPT (Privacy Penetration Test) facility for data owners to carry out independent privacy-pentesting via the TRUMPET platform against their own datasets. The external independent expert will launch attacks against a test dataset in one of the hospitals through the TRUMPET platform. To do this, the expert will be given freedom to use the TRUMPET PPT utility and the Researcher dashboard, having the role of a malicious user of TRUMPET.

3) Characterization of Robustness: Privacy is undoubtedly a critical aspect of the TRUMPET platform, as should always be the case when dealing with sensitive information. However, the robustness of ML-based systems is equally important to establish trust and credibility. To this end, several empirical and formal methods and tools can be used for the test and verification of the desired robustness properties. Some federative platforms [15] can serve as a unique entry point for such an effort, seamlessly handling the inner encoding to a set of state-of-the-art tools, as well as the aggregation of their results and the generation into summarizing tables and graphs. This allows to have a characterization of the ML-based system through pre-defined metrics [16] (*e.g.,* the models are robust within a distance of X around the dataset D) which can support an argumentation of safety to stakeholders and

© 2023 IEEE. Personal use of this material is permitted.
2023 IEEE International Conference on Cyber Security and Resilience (CSR) | DOI: 10.1109/CSR57506.2023.10224961

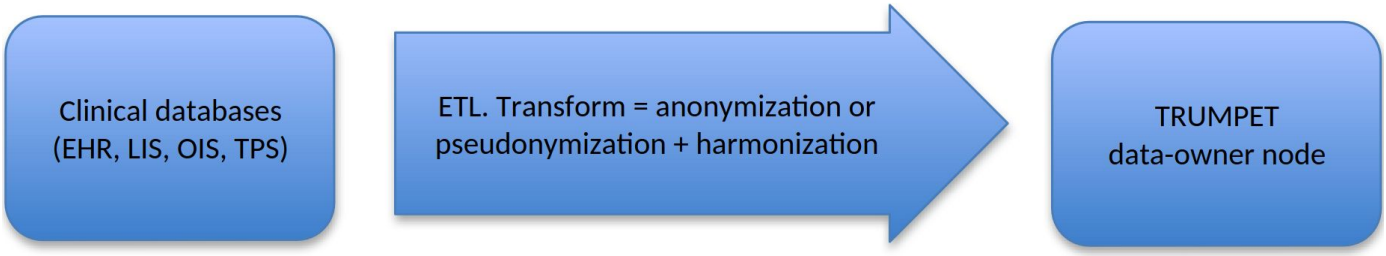


Fig. 4. Process to harmonize clinical data under a common format.

certification bodies. Ultimately, a TRUMPET platform that is both privacy-aware and robust is crucial for building trust, which is essential for its success.

III. Privacy Enhancing Technologies for AFL

The field of Privacy Enhancing Technologies (PETs) has evolved rapidly in recent years, providing increasingly efficient and feasible solutions to the problem of securely processing and sharing sensitive private data. This has resulted in a diverse set of PET flavors [17], where the choice of a specific technique for a particular use case depends on the available resources and the privacy problem to be addressed. Unfortunately, significant obstacles still exist that limit the adoption of PETs in general applications. The obstacles vary from degradation of utility of data, to high computational and communication costs.

The TRUMPET project aims at the application of the most appropiate PET methods for the protection of both exchanged model udpates and the execution of central aggregation. To this end, TRUMPET will study how the performance tradeoffs of baseline PETs improve when tailoring them to the FL setting. This research will take into account several relevant parameters, such as the accuracy/privacy [18] tradeoff, security assumptions [19], computational/communication costs and scalability [20].

The developed methods will be validated in realistic scenarios, thus demonstrating their advantages and limitations and opening new paths for research. Moreover, while many current approaches assume the honest-but-curious attacker threat model, the semi-honest assumption does not always hold in the real world. Beside the expected progress in FL-tailored PETs, we will also improve security by accounting for stronger dishonest parties (aggregators, learning nodes) taking part in FL settings.

A. PET methods in TRUMPET

As it has been already discussed, FL was originally presented as a privacy-preserving technique enabling the training of models on decentralized data sources without needing to transfer the data to a central server. Unfortunately, it has been proven that FL is vulnerable to several inference attacks related to the exposition of the model updates exchanged during training. To ensure resilience to membership inference and attribute leakage attacks, the TRUMPET platform will deploy the most appropriate combination of PETs depending on a number of factors, such as the selected AI model, FL aggregation algorithm, training dataset and the dataset under analysis, as well as scalability and accuracy/performance requirements of the use cases (see Section VI).

The following provides a short description of the most relevant PETs in the context of the TRUMPET project:

1) Homomorphic Encryption (HE): It allows for computation directly on encrypted data [19]. While it provides input privacy for the model updates, it also presents a higher computational cost than other PET techniques. It requires the use of other primitives (e.g., Zero-Knowledge Proofs or ZKPs) to upgrade the security model against stronger adversaries.

2) Secure Multi-Party Computation (SMPC): It allows a group of parties to jointly compute a function while keeping the parties' inputs secret. In general, many SMPC solutions [21], like for example those based on the use of Secret Sharing [22], present a lower computational cost than HE, but usually require a higher number of communication rounds. There is a wide list of different techniques under the label of SMPC (e.g., Secret Sharing, ZKPs, etc).

3) Differential Privacy (DP): It offers statistical privacy guarantees by adding noise to attributes of individual data records before sharing them. While it provides the lowest computational overhead among all the PETs mentioned here, DP [23] entails a tradeoff between privacy level and utility.

4) Coded Distributed Computing (CDC): It is a combination of distributed computing and coding theoretic techniques that enables the distributed computation of a function in the coded domain while keeping the inputs private [24], [25].

IV. A Tool for the Validation of FL Privacy

The TRUMPET project looks at privacy from two complementary points of view. First, we consider statistical privacy as a technological means to avoid information leaking from a computation. Second, we consider privacy from a GDPR point of view. In this last case, we aim at making a connection between the legal GDPR requirements and the statistical requirements, which would help to better fulfill the first type of legal requirements.

A. FL-tailored Privacy Metrics

While Differential Privacy (DP) [26] has become a gold standard notion for statistical privacy in the field of machine

© 2023 IEEE. Personal use of this material is permitted.
2023 IEEE International Conference on Cyber Security and Resilience (CSR) | DOI: 10.1109/CSR57506.2023.10224961

learning, it is a brute force notion. Here, by brute force we mean that it requires to hide secrets with so much noise, that it provides protection under the most reasonable threat model. However, in several use cases, such as in medicine, where precision is important and patient data is expensive, it may not provide the best balance between privacy and utility. Moreover, in FL settings where next to privacy there are security risks, the threat models considered by cryptographic approaches often assume a less strong adversary compared to the one of DP. It is usually good to have all components of a system making consistent and equivalent assumptions.

Frameworks such as the one of Pufferfish privacy [27] offer more opportunity for fine grained modeling, but task the user with deriving all privacy guarantees by himself.

In the TRUMPET project, we will develop statistical privacy metrics: (1) making more fine-grained assumptions, and not only (2) consistent with the other TRUMPET platform components, but also (3) appropriate for our setting in which a limited number of larger, known entities, that can be kept liable for deviation from the protocol, collaborate to train a statistical model on federated data.

In the learning scenario, the proposed privacy metrics will rely on statistically modeling the prior knowledge (or, rather, the uncertainty) that the adversary has regarding the target training datasets. In particular, these metrics will quantify how much knowledge the adversary may gain upon observing the exchanged data in each interaction. This Bayesian approach deviates from the worst-case adversary considered in DP, and is conceptually more aligned with concepts like *distributional privacy* [28] and *Pufferfish privacy* [27]. One advantage of these metrics, besides providing higher utility, is the availability of many information-theoretic results that can be almost directly imported. Another advantage is that its formulation is closer to practical attacks that account for the adversary's uncertainty through sampling a dataset that is assumed to be statistically similar to (i.e. partially overlaps) the target dataset [29], [30].

B. A Tool to Measure TRUMPET Privacy

Having a definition of a privacy metric does not lead easily to being able to actually measure privacy according to that metric. For example, there is a vast literature considering specific algorithms in isolation and studying their privacy guarantees, but far less attention has been paid to the setting where a researcher interacts with a federated data set and decides on the next query after having seen the answer to the previous one. In order to address this issue, TRUMPET will develop a tool that keeps track of the privacy budget a researcher has used so far. By doing so, we aim at not only using the classic composition rules, which have been studied for differential privacy, but also to let the tool look more intelligently at the queries, searching for tighter upper bounds of the actual privacy budget consumed.

In the FL scenario, it is often the case that the Data Owners also interact with the Aggregator through the exchange of updates of the model that is being learned. Each of these updates potentially leaks information regarding the contents of the training dataset available at each Data Owner. The availability of a privacy budget and, more importantly, a practical way of measuring how much information has been leaked becomes crucial in such an interactive scenario: on the one hand, it may serve to establish to which extent the exchanged data must be protected/obfuscated by PET methods; on the other hand, it may lead to hard rules on the federation, for example, by preventing a Data Owner from submitting further updates whenever the budget limit has been reached.

V. GDPR Compliance of AFL

Regulation (EU) 2016/679 (General Data Protection Regulation, GDPR) applies to any processing of personal data, including the training of machine learning (ML) algorithms on data relating to identifiable individuals. Federated learning (FL) has gained recent popularity as a privacy-preserving technology that can help ML solutions achieve GDPR compliance due to its capability to train algorithms on various datasets without the transfer of any personal data. As data never leaves its source and is not collected in a central location, FL solutions can be used to ensure better security for personal data. This was recently highlighted by ENISA, which confirmed that FL can help preserve the privacy of data and protect against the unauthorized disclosure of sensitive data [31]. In addition, FL architectures are considered to facilitate compliance by design with the data protection principles of data minimization and storage limitation [32], [33]. In other words, they can help ensure that only relevant, adequate and necessary personal data is processed, and kept for as little time as possible.

FL can therefore be used as a technical measure to achieve compliance with the principles of data security and data minimization. The use of FL alone, however, is not sufficient to ensure that the data processing is GDPR compliant. First of all, FL can be susceptible to a number of privacy and security challenges, including man-in-the-middle and insider attacks, and most importantly, inference attacks that aim to re-identify data subjects from FL's AI model parameter updates. The AFL methods TRUMPET will develop combining FL with other PETs are therefore crucial for reducing the risk of re-identification of data subjects (see Sections III and IV).

Second, while AFL methods are essential for ensuring data privacy and security, data controllers using the TRUMPET AFL platform remain responsible for ensuring compliance with all GDPR principles. In particular, they should implement measures to prevent any unfair or arbitrary treatment of data subjects, including by performing Data Protection Impact Assessments (DPIAs) to assess any risks for the data subjects. Data owners should ensure that in cases where previously collected personal data are re-used for training ML models, the new purpose for which they are processing those data is compatible with the original one. Additionally, they are responsible for determining the lawful basis for secondary data processing and demonstrating compliance with all GDPR principles and requirements.

© 2023 IEEE. Personal use of this material is permitted.
2023 IEEE International Conference on Cyber Security and Resilience (CSR) | DOI: 10.1109/CSR57506.2023.10224961

VI. Research with Distributed Clinical Data

The AFL-based TRUMPET platform will give the possibility for researchers to develop statistical AI models from distributed clinical datasets, while providing strong privacy guarantees and without transferring data out of each hospital's infrastructure. We will consider three different clinical use cases to demonstrate the capability of the TRUMPET platform in the field of cancer treatment and diagnosis.

A. Use Cases' Description

1) Non-small Cell Lung Cancer (NSCLC) Use Case: It is the primary cause of cancer-related death worldwide. Patients with advanced non-oncogene addicted disease usually benefit from treatment with immune checkpoint inhibitors (ICIs), which are able to reactivate the repressed immune response against tumor cells. However, only a portion of patients achieves durable clinical benefit from these treatment approaches [34].

2) Stereotactic Body Radiation Therapy (SBRT) Use Case: For some cancers, surgery remains a mainstay in the treatment of solitary metastases. However, for patients with metastasis unresectable by surgery, SBRT is increasingly used as an ablative treatment option. Although at present there are several studies investigating the utility of SBRT in the treatment of metastatic disease, currently, limited guidelines allow to determine which patients could really benefit from it [35].

3) Head and Neck Cancer (HNC) Use Case: Despite the therapeutic goals of RT (Radiation Therapy) as a potentially curative treatment in HNC, toxicity is commonly seen. Patients undergoing radiation therapy for HNC can experience significant early and long-term side effects. We hypothesize that a systemic assessment of late side effects against planned dose-volume histograms and RT fractions regimen could generate new knowledge to further optimize treatment planning in HNC [36].

B. Secondary TRUMPET objectives

In particular, several secondary TRUMPET objectives have been established in relation to each of these use cases:

- **Non-small cell lung cancer (NSCLC) use case**: We intend to integrate clinical, biological and radiological data of NSCLC patients treated with immunotherapy to find an algorithm predictive of patient's prognosis.
- **Stereotactic body radiotherapy (SBRT) use case**: We intend to develop a classifier predicting the survival probability over 6 months, hence obtaining one criteria for the eligibility to SBRT treatment. Specifically, the model should answer the two following questions:
  - What is the probability for a patient X to survive more than 4 or 6 months?
  - What is the estimated survival period of patient X?
- **Head and neck cancer (HNC) use case**: We intend to identify causality relationships between the radiotherapy treatment plan and its delivery towards late side effects in Head and Neck cancer.

C. Ethical Aspects of the Clinical Use Cases

The main ethical aspects of the TRUMPET project relate to the use of retrospective patient health data for the three TRUMPET platform use cases. The required ethical approval actions for the use of patient health data are being prepared and performed by IRST IRCCS and CHU de Liège before starting work on any of the above described use cases. In order to address and mitigate potential risks for data subjects stemming from the secondary use of data within TRUMPET, a Data Protection Impact Assessment (DPIA) will be performed, and the necessary technical and organisational measures (such as de-identification and encryption of personal data) will be implemented throughout the project to ensure the security of patient data and compliance with data protection requirements. See Section V for more details on the compliance of the TRUMPET platform with GDPR principles.

VII. Some conclusions and expected results

This work has given an overview of the scope and main technical aspects of the EU-funded project TRUMPET:

- The main objective is to provide a highly scalable FL-based service platform for researchers.
- The TRUMPET platform will allow to perform collaborative studies with cross-border European datasets and with privacy guarantees that follow GDPR requirements.
- Two concrete technical objectives: (1) research and development of FL-tailored PET methods, and (2) research and development of a set of novel privacy metrics and a tool for the privacy evaluation of FL implementations.
- Identify and perform a legal study of the implications of GDPR in FL implementations. Moreover, even if the GDPR is an European regulation, different European countries may apply further restrictions. TRUMPET will necessarily take this into account in the legal study.
- The project will engage external penetration test experts to validate the privacy protection, and provide a facility for data owners to do independent privacy-pentesting.
- The use cases cover major unmet clinical needs in treatment and diagnosis of cancer. They emphasize different points of view over the cohorts, either from a disease perspective or from a treatment modality perspective. Thanks to the federated nature of the platform and its privacy metric, the pilots will demonstrate the benefits of Armored FL to streamlining clinical and epidemiological research processes across the EU, while ensuring compliance to the GDPR and national data privacy laws.

Acknowledgment

Trumpet project was funded by the European Union with grant agreement Nr.101070038.

Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

© 2023 IEEE. Personal use of this material is permitted.

2023 IEEE International Conference on Cyber Security and Resilience (CSR) | DOI: 10.1109/CSR57506.2023.10224961

REFERENCES

[1] "TRUMPET project," https://trumpetproject.eu/, 2023, [Online; accessed 31-March-2023].

[2] "CORDIS webpage: TRUMPET project," https://cordis.europa.eu/project/id/101070038, 2022, [Online; accessed 31-March-2023].

[3] M. Khan, F. G. Glavin, and M. Nickles, "Federated learning as a privacy solution - an overview," *Procedia Computer Science*, vol. 217, pp. 316–325, 2023, 4th International Conference on Industry 4.0 and Smart Manufacturing.

[4] T. Yang, G. Andrew, H. Eichner, H. Sun, W. Li, N. Kong, D. Ramage, and F. Beaufays, "Applied federated learning: Improving google keyboard query suggestions," *CoRR*, vol. abs/1812.02903, 2018. [Online]. Available: http://arxiv.org/abs/1812.02903

[5] D. Leroy, A. Coucke, T. Lavril, T. Gisselbrecht, and J. Dureau, "Federated learning for keyword spotting," in *IEEE ICASSP*. IEEE, 2019, pp. 6341–6345.

[6] J. Ogier du Terrail, A. Leopold, C. Joly, C. Béguier, M. Andreux, C. Maussion, B. Schmauch, E. W. Tramel, E. Bendjebbar, M. Zaslavskiy, G. Wainrib, M. Milder, J. Gervasoni, J. Guerin, T. Durand, A. Livartowski, K. Moutet, C. Gautier, I. Djafar, A.-L. Moisson, C. Marini, M. Galtier, F. Balazard, R. Dubois, J. Moreira, A. Simon, D. Drubay, M. Lacroix-Triki, C. Franchet, G. Bataillon, and P.-E. Heudel, "Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer," *Nature Medicine*, vol. 29, no. 3–4, pp. 135–146, 2023.

[7] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. A. Bonawitz, Z. Charles, G. Cormode, R. Cummings, R. G. L. D'Oliveira, H. Eichner, S. E. Rouayheb, D. Evans, J. Gardner, Z. Garrett, A. Gascón, B. Ghazi, P. B. Gibbons, M. Gruteser, Z. Harchaoui, C. He, L. He, Z. Huo, B. Hutchinson, J. Hsu, M. Jaggi, T. Javidi, G. Joshi, M. Khodak, J. Konečný, A. Korolova, F. Koushanfar, S. Koyejo, T. Lepoint, Y. Liu, P. Mittal, M. Mohri, R. Nock, A. Özgür, R. Pagh, H. Qi, D. Ramage, R. Raskar, M. Raykova, D. Song, W. Song, S. U. Stich, Z. Sun, A. T. Suresh, F. Tramèr, P. Vepakomma, J. Wang, L. Xiong, Z. Xu, Q. Yang, F. X. Yu, H. Yu, and S. Zhao, "Advances and open problems in federated learning," *Found. Trends Mach. Learn.*, vol. 14, no. 1-2, pp. 1–210, 2021.

[8] P. Vanhaesebrouck, A. Bellet, and M. Tommasi, "Decentralized collaborative learning of personalized models over networks," in *AISTATS*, ser. Proceedings of Machine Learning Research, vol. 54. PMLR, 2017, pp. 509–517.

[9] H. Tang, X. Lian, M. Yan, C. Zhang, and J. Liu, "D^2: Decentralized training over decentralized data," in *ICML*, ser. Proceedings of Machine Learning Research, vol. 80. PMLR, 2018, pp. 4855–4863.

[10] B. E. Woodworth, J. Wang, A. D. Smith, B. McMahan, and N. Srebro, "Graph oracle models, lower bounds, and gaps for parallel stochastic optimization," in *NeurIPS 2018*, 2018, pp. 8505–8515.

[11] L. Melis, C. Song, E. D. Cristofaro, and V. Shmatikov, "Exploiting unintended feature leakage in collaborative learning," in *IEEE Symposium on Security and Privacy*. IEEE, 2019, pp. 691–706.

[12] M. Mansouri, M. Önen, W. B. Jaballah, and M. Conti, "Sok: Secure aggregation based on cryptographic schemes for federated learning," *Proc. Priv. Enhancing Technol.*, vol. 2023, no. 1, pp. 140–157, 2023.

[13] P. Blanchard, E. M. E. Mhamdi, R. Guerraoui, and J. Stainer, "Machine learning with adversaries: Byzantine tolerant gradient descent," in *NIPS*, 2017, pp. 119–129.

[14] M. Nasr, R. Shokri, and A. Houmansadr, "Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning," in *IEEE Symposium on Security and Privacy*. IEEE, 2019, pp. 739–753.

[15] J. Girard-Satabin, M. Alberti, F. Bobot, Z. Chihani, and A. Lemesle, "Caisar: A platform for characterizing artificial intelligence safety and robustness." To appear in the proceedings of IJCAI's AISafety workshop, 2022. [Online]. Available: https://arxiv.org/abs/2206.03044

[16] J. Mattioli, S. Henri, D. Agnes, A.-F. Kahina, A. Afef, C. Zakaria, S. Khalfaoui, and G. Pedroza, "An overview of key trustworthiness attributes and kpis for trusted ml-based systems engineering," *AI Trustworthiness Assessment*, 2023.

[17] Big Data UN Global Working Group, "UN Handbook on Privacy-Preserving Computation Techniques," https://unstats.un.org/bigdata/task-teams/privacy/index.cshtml, 2022, [Online; accessed 31-March-2023].

[18] R. C. Geyer, T. Klein, and M. Nabi, "Differentially private federated learning: A client level perspective," *CoRR*, vol. abs/1712.07557, 2017. [Online]. Available: http://arxiv.org/abs/1712.07557

[19] M. Chase, H. Chen, J. Ding, S. Goldwasser, S. Gorbunov, J. Hoffstein, K. Lauter, S. Lokam, D. Moody, T. Morrison, A. Sahai, and V. Vaikuntanathan, "Security of homomorphic encryption," HomomorphicEncryption.org, Redmond, WA, Tech. Rep., July 2017.

[20] J. Cabrero-Holgueras and S. Pastrana, "Sok: Privacy-preserving computation techniques for deep learning," *Proc. Priv. Enhancing Technol.*, vol. 2021, no. 4, pp. 139–162, 2021.

[21] D. Evans, V. Kolesnikov, and M. Rosulek, "A pragmatic introduction to secure multi-party computation," *Foundations and Trends® in Privacy and Security*, vol. 2, pp. 70–246, 2018.

[22] I. Damgård, M. Keller, E. Larraia, V. Pastro, P. Scholl, and N. P. Smart, "Practical covertly secure MPC for dishonest majority - or: Breaking the SPDZ limits," in *ESORICS*, ser. Lecture Notes in Computer Science, vol. 8134. Springer, 2013, pp. 1–18.

[23] C. Dwork, "Differential privacy," in *International Colloquium on Automata, Languages, and Programming*. Berlin, Heidelberg: Springer, 2006, pp. 1–12.

[24] J. S. Ng, W. Y. B. Lim, N. C. Luong, Z. Xiong, A. Asheralieva, D. Niyato, C. Leung, and C. Miao, "A survey of coded distributed computing," vol. abs/2008.09048, 2020. [Online]. Available: https://arxiv.org/abs/2008.09048

[25] S. Ulukus, S. Avestimehr, M. Gastpar, S. Jafar, R. Tandon, and C. Tian, "Private retrieval, computing and learning: Recent progress and future challenges," vol. abs/2108.00026, 2021. [Online]. Available: https://arxiv.org/abs/2108.00026

[26] C. Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 1–277, 2014.

[27] D. Kifer and A. Machanavajjhala, "Pufferfish: A framework for mathematical privacy definitions," *ACM Transactions on Database Systems*, vol. 39, no. 1, pp. 3:1–3:36, 2014.

[28] A. Blum, K. Ligett, and A. Roth, "A learning theory approach to non-interactive database privacy," in *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*. ACM, 2008, pp. 609–618.

[29] M. Nasr, R. Shokri, and A. Houmansadr, "Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning," in *2019 IEEE Symposium on Security and Privacy, SP 2019, San Francisco, CA, USA, May 19-23, 2019*. IEEE, 2019, pp. 739–753.

[30] N. Carlini, S. Chien, M. Nasr, S. Song, A. Terzis, and F. Tramèr, "Membership inference attacks from first principles," in *43rd IEEE Symposium on Security and Privacy, SP 2022, San Francisco, CA, USA, May 22-26, 2022*. IEEE, 2022, pp. 1897–1914.

[31] ENISA, "Securing Machine Learning Algorithms," https://www.enisa.europa.eu/publications/securing-machine-learning-algorithms, 2021, [Online; accessed 31-March-2023].

[32] S. Rossello, R. D. Morales, and L. Muñoz-González, "Data Protection by design in AI?" *Computerrecht: tijdschrift voor informatica en recht*, pp. 1–11, 2021.

[33] Norwegian Data Protection Authority, "Report on Artificial intelligence and privacy," https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf, 2018, [Online; accessed 31-March-2023].

[34] R. S. Vanguri, J. Luo, A. T. Aukerman, J. V. Egger, C. J. Fong, N. Horvat, A. Pagano, J. d. A. B. Araujo-Filho, L. Geneslaw, H. Rizvi, R. Sosa, K. M. Boehm, S.-R. Yang, F. M. Bodd, K. Ventura, T. J. Hollmann, M. S. Ginsberg, J. Gao, R. Vanguri, M. D. Hellmann, J. L. Sauter, S. P. Shah, and M. M. Consortium, "Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer," *Nature Cancer*, vol. 3, no. 10, pp. 1151–1164, 2022.

[35] S. Ramadan, K. Quan, K. Schnarr, R. A. Juergens, S. J. Hotte, S. D. Mukherjee, A. Kapoor, B. M. Meyers, and A. Swaminath, "Impact of stereotactic body radiotherapy (sbrt) in oligoprogressive metastatic disease," *Acta Oncologica*, vol. 61, no. 6, pp. 705–713, 2022.

[36] F. Siddiqui and B. Movsas, "Management of radiation toxicity in head and neck cancers," *Seminars in Radiation Oncology*, vol. 27, no. 4, pp. 340–349, 2017.

© 2023 IEEE. Personal use of this material is permitted.

2023 IEEE International Conference on Cyber Security and Resilience (CSR) | DOI: 10.1109/CSR57506.2023.10224961