NeurIPS 2023 Competition: Privacy Preserving Federated Learning Document VQA

Dimosthenis Karatzas[*] Rubèn Tito[*] Mohamed Ali Souibgui[*] Khanh Nguyen[*]
Raouf Kerkouche[†] Kangsoo Jung[‡] Marlon Tobaben[§] Joonas Jälkö[§]
Vincent Poulain[¶] Aurelie Joseph[¶] Ernest Valveny[*] Josep Lladós[*] Antti Honkela[§]
Mario Fritz[†]

dimos@cvc.uab.es

Abstract

The Privacy Preserving Federated Learning Document VQA (PFL-DocVQA) competition aims to challenge the community with developing provable private and communications efficient solutions in a federated setting, for a real-life use case: invoice processing.

The competition puts forward a dataset of real invoice documents, and associated questions and answers that require information extraction and reasoning over the document images. The objective of the competition participants would be to fine tune a pre-trained, generic, state of the art Document Visual Question Answering model provided by the organisers on the new domain. The training will take place in a federated setting resembling a typical invoice processing setup. The base model is a multi-modal generative language model, and the sensitive information might be exposed through the visual and/or the textual input modality.

The PFL-DocVQA competition will run in two stages following a "Blue team / Red team" scheme. In the first stage provable privacy preserving, federated learning solutions will be solicited, while in the second stage membership inference attacks will be designed against the top performing methods. We propose to run the first stage in the timeline of NeurIPS 2023.

We envisage that the competition will provide a new testbed for developing and testing private federated learning methods, while at the same time it is expected to raise awareness about privacy in the document image analysis and recognition community.

Keywords

Differential privacy, Federated Learning, Document Understanding, Document Visual Question Answering

[*]Computer Vision Center, Universitat Autonoma de Barcelona, Spain
[†]CISPA Helmholtz Center for Information Security, Germany
[‡]French Institute for Research in Computer Science and Automation (INRIA), France
[§]University of Helsinki, Finland
[¶]Yooz, France

Preprint. Under review.

1 Competition description

1.1 Background and impact

The objective of the Privacy Preserving Federated Learning Document VQA (PFL-DocVQA) competition is to develop privacy-preserving solutions for fine-tuning multi-modal language models for document understanding. Specifically, we will seek efficient federated learning solutions for finetuning a pre-trained generic Document Visual Question Answering (DocVQA) model on a new domain, that of invoice processing. The submitted solutions should be accompanied by proofs that ensure that sensitive information in the training set remains confidential through the training process.

The competition, which is organised in the context of the European Lighthouse on Safe and Secure AI, will comprise two stages in a "Blue team / Red team" scheme. In the first stage (Blue team), privacy preserving federated learning solutions will be solicited and will be measured in terms of performance on the task of DocVQA, and efficiency in terms of federated learning communications costs. The solutions that provide provable privacy guarantees will be further reviewed and go through a second stage (Red team), where different teams will design membership inference attacks against these methods. We propose to run the first stage in the timeline of NeurIPS 2023, while the second stage will take place later, possibly linked with NeurIPS 2024.

The topic of Document Understanding and specifically Document VQA was chosen on one hand due to the important scientific and technical challenges it presents, and on the other hand because it is a real-life scenario, well adapted to data federation, with clear socio-economic importance.

Multi-modal language models (transformer-based, generative, vision and language models) are typically used for DocVQA, frequently built on backbones pre-trained on unavailable large scale data. This raises numerous research questions about the extent to which such models can be fine-tuned in an efficient privacy-preserving manner on downstream tasks, and to the extent that sensitive information exposed through the visual and/or the textual input modality can be protected. Technically, these models are orders of magnitude larger than typical models used in federated learning research, calling for more efficient methods.

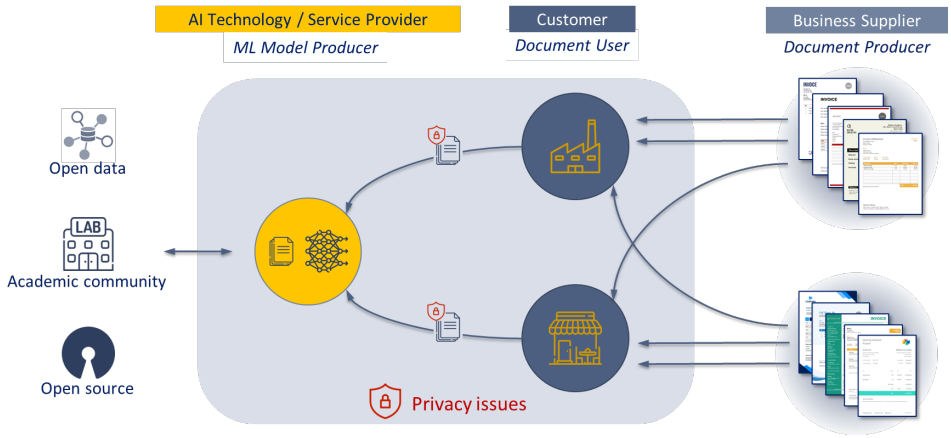


Figure 1: Overview of the Invoicing scenario. Suppliers generate invoices for their customers. To process the invoice, the customer applies Document Intelligence technologies. Privacy issues raise in all communications of documents between document users (customers) and AI service providers.

Motivation. This competition reproduces a typical real-life invoicing scenario, where business suppliers produce invoices for their services and send them to their customers. These documents contain sensitive information (who consumed/purchased, what, why, when, how much did they pay, how much a particular supplier charged for a service, which account to make a payment in etc.) and the customers (document users) need to extract this information and take the corresponding actions (i.e. reject, or make a payment against the invoice). In automated pipelines, these documents would be sent to AI technology providers, typically offered in the form of cloud services [6] , which

[6] Automatic document processing services offered by large corporates (AWS Intelligent Document Processing, Google Cloud Document AI, Microsoft Azure Form Recognizer, etc) or specialised providers.

automatically extract all required information from the documents, and return it to the document users. This information is then used to effectuate the corresponding bank transactions.

Training an accurate DocVQA model requires a considerable amount of data that one entity (also known as client) may not have. One possible solution to this problem is to train this model collaboratively by aggregating and centralizing data from a set of clients that face the same problem. Unfortunately, documents can be sensitive and can leak valuable information.

Another solution is to consider Federated Learning (FL), which allows collaborative model training without sharing the entire data set of a client. Instead, only the trained model and its updates are shared between a central server and the different clients. However, even though federated learning is more private than the centralized approach, many attacks have shown that a significant amount of information can still be inferred from the updates/models shared between the clients and the server during training or afterwards. Adversaries can be either a participant, the server, or an external entity with access to the released trained model. One known type of attack is to reconstruct the records included in the dataset of a client (26; 25; 2; 24; 12; 4; 16; 13). Another type of attack is to infer whether a specific record is included in the dataset of a client, which is called "membership inference attack" (17; 16). Finally, a property inference attack (16) is used to determine whether a property about a group of people is included in the client's dataset.

Differential Privacy (DP), which is considered the gold standard in terms of privacy preservation, can be used to provide theoretical privacy guarantees. However, differential privacy introduces a trade-off between utility and privacy. Specifically, differential privacy requires the addition of random noise during training, which has an impact on the accuracy of the model. This noise increases with the desired level of privacy, but is inversely proportional to the accuracy of the model.

Another drawback of federated learning is the high communication cost. Indeed, at each federated round, the global model is sent by the server to selected entities (downstream step) to train it on their local data and then the update of this model is sent by these selected entities to the server (upstream step). Knowing that the model is composed of millions or even billions of parameters for the most recent architectures and that it takes more than one federated round for the model to converge and reach a good accuracy, this requires a significant communication cost.

This competition is designed to expose the above challenges and invite the community to design novel creative solutions for this real-life use case.

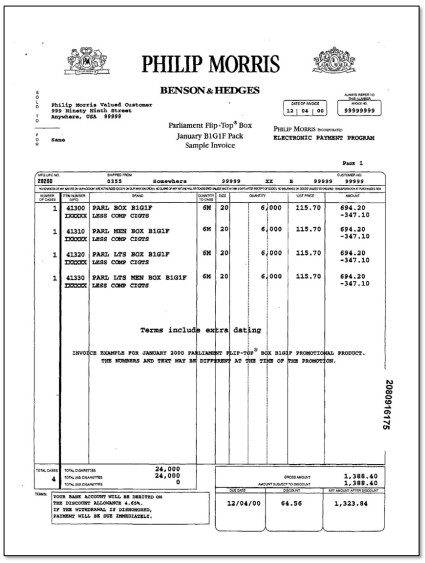
Scientific Impact. The scientific impact of this competition is two-fold. On one hand it is expected that it will provide a new testbed for advancing research in differential privacy and federated learning in a multi-modal scenario. The focus on DocVQA allows the community to concentrate efforts on much larger models than the typical research scenarios to date. On the other hand, it is expected to bring closer the document image analysis and recognition (DIAR) and the privacy community, raising privacy awareness in the DIAR community.

Socio-economic Impact: During 2019, around 550 billion invoices were used globally and 90% of them were exchanged as document papers/images, this amount of invoices is expected to quadruple in size by 2035 (11). Document capture and management software are key elements across a wide variety of economic sectors: public administration, banking, energy, financial services, insurance, healthcare, retail, telecom & IT, transportation and logistics, etc. The global Document Capture Software Market size was estimated at USD $2,153.68$ million in 2021, USD $2,382.40$ million in 2022, and is projected to grow at a CAGR of 10.79% to reach USD $3,984.29$ million by 2027 (18). Despite the high volume of documents being processed, and the potentially sensitive information they contain, the application of privacy-protecting methods has not yet been explored at the model level.

Relevance and expected participants: The main target for this competition are the Differential Privacy and Federated Learning communities, to whom we will bring a new and realistic multi-modal scenario. The competition is expected to also generate interest in the Document Image Analysis and Recognition community, as it addresses the core problem of document privacy in state of the art DocVQA methods.

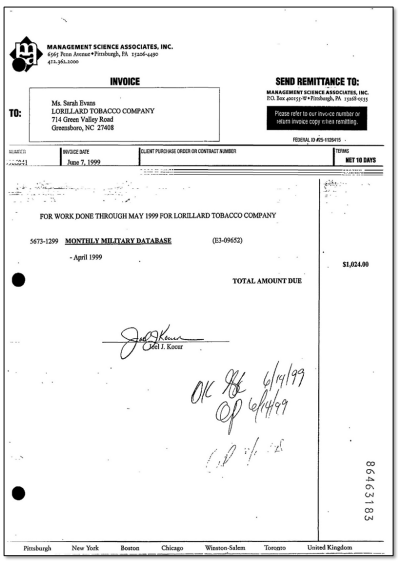
1.2 Novelty

The proposed competition is novel and not connected to any previous series. It reuses and significantly extends the DocILE dataset (see next section).



Q: Is the amount due more than $2,000.00?
A: No

Q: What are the payment terms?
A: Pay immediately



Q: What is the invoice date?
A: June 7,1999

Q: Is this invoice from the Management Science Associates, INC.
A:Yes

Figure 2: Example documents of PFL-DocVQA dataset with its corresponding questions.

In the federated learning and privacy domain, the PETs Prize Challenge [7] is a recent, similarly organised competition, featuring a Blue Team / Read Team scheme. Contrary to PFL-DocVQA, the tasks of the PETs Prize challenge are based on tabular data (Financial Crime and Pandemic Forecasting) and baselines are built on classical ML algorithms. The proposed competition focuses on real, multi-modal data and large deep learning models.

The organizing team has organised successfully a series of competitions [8] on DocVQA (15; 20). These previous competitions were organised on different datasets, and aimed on non-private models. In contrast, PFL-DocVQA aims to develop efficient and private training techniques for a representative state of the art model.

1.3 Data

The PFL-DocVQA dataset is created by defining question and answer pairs on documents from the public DocILE (19) dataset. The original DocILE dataset is designed for Key Information Localization and Extraction (KILE) on invoice-like documents. DocILE is composed of $6,680$ annotated and $932K$ non-annotated real business documents from publicly available sources.

The questions framed on the DocILE annotated documents are produced using template questions on the existing key-value pairs. Afterwards, the questions are paraphrased using automatic tools like NLTK WordNet to add linguistic variability. In addition, we are adding key-value annotations to a significant portion of the $932K$ non-annotated invoices in the dataset, using professional document analysis software from our co-organiser Yooz. The automatically obtained predictions will go through a manual verification stage. We expect to annotate an extra $15,000$ documents in addition to the original $6,680$ documents.

The sensitive information that needs to be kept private is the providers' ID. For the Blue team stage, the distribution of data into federated learning nodes will reflect a realistic scenario, where different nodes receive invoices from disjointed sets of providers, resulting in a highly non i.i.d. distribution.

[7] https://www.drivendata.org/competitions/98/nist-federated-learning-1/
[8] https://rrc.cvc.uab.es/?ch=17&com=introduction

In the current version of DocILE there are $1,890$ providers with an unbalanced number of invoices contributed per provider. This is expected to rise to approximately $10K$ different providers after the extended annotation process.

The DocILE documents are published under the MIT License. New annotations and Question-Answer pairs provided by us will be also made freely available.

1.4 Privacy Model

We consider an adversary, or a set of colluding adversaries, who can access the released trained model after the federated learning process. The adversary aims at inferring the training data (or some private information about them) of the participating clients' groups. The adversary is passive (i.e., honest-but-curious), that is, it does not modify the trained global model.

Different privacy requirements are considered depending on the information the adversary seeks to infer. In general, private information can be inferred about:

- any record (user) in any dataset of any client (record-level privacy) (23; 9).
- one or more groups of records in any dataset of any client (group-level privacy) (14; 3),
- any client/party (client-level privacy) (5; 10; 8).

For our challenge, we seek to protect the identity of providers that could be exposed to textual (company name, account number) or visual (logo, presentation) information. If a malicious competitor (adversary) manages to infer information about a company's providers, it could have a direct impact on the company's business.

Several invoices in a client's dataset may come from the same provider and therefore represent a group associated with this provider. Therefore, differential privacy at the group level is the most appropriate level of protection in this case.

Therefore, the adversary should not be able to learn from the received model or its updates whether any group's data is involved in the federated run (up to ε and δ). This adversarial model is appropriate in many other practical applications when the confidential information spans over multiple samples in the training data of a single client (e.g., the presence of a group of samples, such as people from a certain race). Group-level differential privacy guarantees plausible deniability to any group of samples inside any client's dataset. Therefore, any negative privacy impact on a group cannot be attributed to their involvement in the protocol run.

1.5 Tasks and application scenarios

1.5.1 Tasks set-up

This competition comprises two stages that will run sequentially, out of which we propose to run the first stage (Blue Team) aligned to the timeline of NeurIPS 2023. We include also some brief information about the second stage (Red Team) to better contextualize the overall competition scheme.

Blue Team. The participating teams of the first stage will create methods to train Document Visual Question Answering models on the provided documents with privacy guarantees, using federation. For clarity, we name the participating teams of this stage as Blue Teams. The stage will comprise two tracks. In **Track 1**, the methods will be trained within a federated learning framework, simulating the need for cooperation between different entities to achieve the best performing model in the most efficient way. In **Track 2**, in addition to the use of federated learning, the participants will be required to apply differential privacy algorithms that must comply with a fixed privacy budget. At the end of this stage, the Blue Teams will submit their models, weights, and technical reports describing the models. Track 2 participants should include the theoretical privacy proof implemented by their methods in the reports.

Red Team. During the second stage, the participating teams (Red Teams) will be given access to the Blue Teams' methods submitted to Track 2, along to their privacy proofs, and their objective will be to device membership inference attacks indicating if data from any of the providers in the Expanded Training set (see figure 3) has been used during the training process or not.

Teams can participate in any or both of the stages. The only restriction is that a red team cannot perform a membership attack over its own submitted methods.

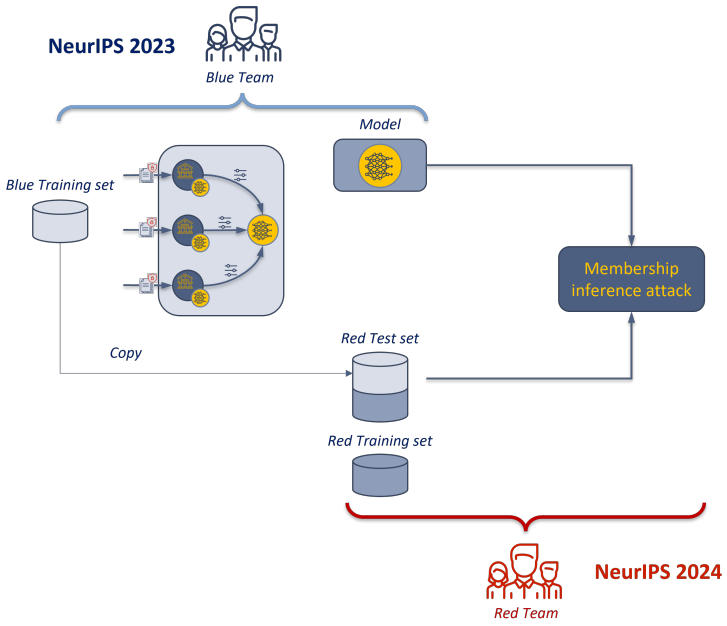


Figure 3: The competition will follow a Blue team / Red team scheme. The Blue team will produce privacy-preserving models and corresponding theoretical guarantees, while the Red team will device attacks aiming to refute the original claims.

1.5.2 NeurIPS relevance

This competition is relevant to NeurIPS, primarily to the "privacy and security", and secondarily to the "computer vision" and "NLP" areas. We consider that NeurIPS provides a suitable setting to hold this competition as it targets the Federated Learning and Differential Privacy communities, while it is expected to attract the attention of the Document Image Analysis and Recognition community.

1.6 Metrics

The trained model produced by the Blue teams will be evaluated in terms of two utility metrics: (1) performance in the VQA task and (2) performance in terms of communication efficiency.

The Red Team -which will not be part of this competition- will perform a membership inference attack, indicating which of the providers with documents in the Red Training set (see figure 3) was used during the training process. We will measure the accuracy of these predictions.

1.7 Baselines, code, and material provided

We will provide two types of baselines. First, a non-private, centrally trained version of the proposed document visual question answering model (21) that will indicate a theoretical upper bound in terms of VQA performance. Second, a basic deferentially private, federated learning training method, on which participants can build their own variants.

We will provide a *"starting kit"* at the beginning of the competition, following the calendar detailed in section 2.3. This will consist of the non-private and the differentially private federated learning baseline implementations including sample code to measure communication efficiency, and the script needed to calculate the baselines' ε budget for some given parameters. The code will be made publicly available through GitHub.

- **Question-answering performance:** We will use two typical DocVQA metrics (15; 22) to evaluate the performance of the VQA task: Accuracy and Average Normalised Levenshtein Similarity (ANLS). Accuracy is a hard metric measuring the percentage of times the model produces the correct answer, while ANLS is a soft counterpart, more graceful to OCR errors.

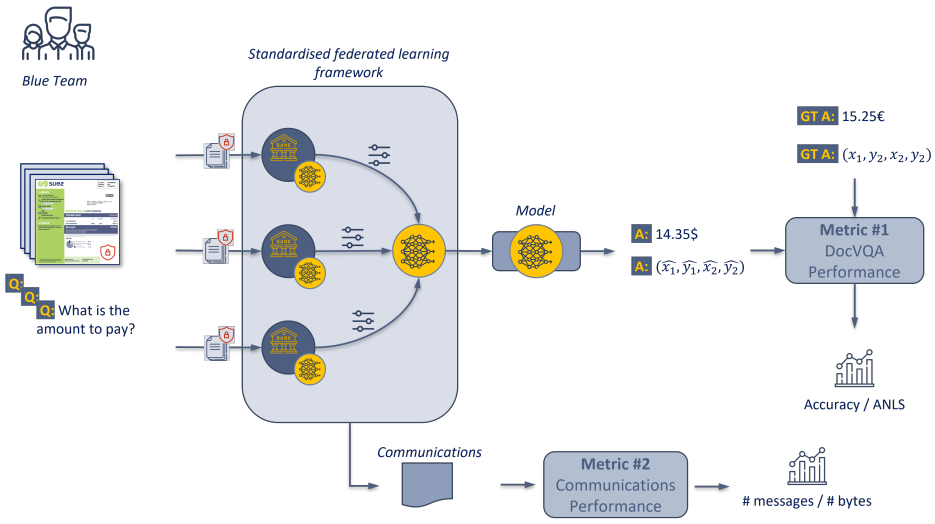


Figure 4: The Blue team will have to devise models and give corresponding privacy guarantees. A standardised federated learning framework will be used to report the amount of communications used. Final models will be evaluated in terms of their utility, with standard DocVQA performance metrics

- **Communication efficiency:** We will measure the amount of data communicated during training in terms of bytes transferred during the federated learning cycles. We will provide example code to log this information for popular federated learning frameworks.

1.8 Website, tutorial and documentation

The competition will be hosted on the ELSA benchmarks platform. The portal shares the same codebase as the Robust Reading Competition portal, also implemented by the organisers, which serves >41k researchers.The competition Web will include detailed information about task participation, links to resources provided, contact details and an automatically updated leaderboard that will be made public at the end of the competition (to avoid cherry picking). This will be complemented by a GitHub repository that will include detailed instructions on using the baselines.

2 Organizational aspects

2.1 Protocol

Registration and submission steps: Participants will be required to register to the ELSA benchmarks portal in order to get access to the datasets and submit results. At registration time, they will have to agree to comply with the competition rules. Then, they will download the dataset, train their models offline, and upload the requested results, which include (1) the trained models (weights and code) that allows to reproduce the experiments and run the model over the test set (2) the communication efficiency in terms of the number of bytes communicated in each of the FL rounds, and (3) for Track 2 participants, a report with the description of the differential privacy algorithm and its theoretical proof. All the evaluation will take place automatically on the ELSA benchmarks platform.

Cheating and overfitting prevention:

- The test set will not be released during the competition. Instead, the organising team will run the submitted models on it.
- No variants of the same method (with different hyperparameters) will be permitted.
- The organizing team will review the differential privacy algorithms' description and theoretical proof of the top-scoring methods.
- The organizers will run the federated learning framework with the code of the top-scoring methods to ensure that the communications log is correct.

2.2 Rules and Engagement

Rules: All the participants must comply with the following rules:

- Participants can only use the provided DocVQA model to design their training methods, and the methods must be fine tuned on the provided dataset only.
- Participants will be required to provide the model code and weights to the organisers, so it can be independently tested.
- Methods must be trained adhering to the provided federated learning set-up.
- Participants will be required to report the amount of communicated information during the FL training as computed by the provided function.
- Methods must have a privacy budget of no more than a pre-defined (ε, δ) budget.
- Participants must submit a report describing their training algorithm along with a theoretical privacy proof.

Communication: Participants will be able to communicate with the organizers through the competition email. On the other hand, the competition platform features a public *Challenge News* section. Finally, for urgent or important communications, the organizers can communicate with the participants through the provided email used to register to the competition.

2.3 Schedule and readiness

This is the tentative schedule for the competition:

- June 1, 2023: Competition registration opens.
- June 30, 2023: Training and validation sets released. Baselines released.
- Oct. 27, 2023: End of the competition. Code submission deadline.
- Nov. 1, 2023: Privacy proof reports due for track 2 participants.
- Nov. 15, 2023: Winning teams announced
- Dec. 10-16, 2023: Results presented at NeurIPS

2.4 Competition promotion and incentives

We plan to distribute the call to research mailing lists (ML, CV, IAPR TC10 and TC11, PET symposium, and TrustworthyML list), and social media , as well as the ELSA project communication channels to promote participation to the competition.

3 Resources

3.1 Organizing team

This competition is organised in the context of the European Lighthouse on Secure and Safe AI (ELSA) project, which leads research efforts in three important areas of Artificial Intelligence: technical robustness and safety, privacy, and human agency and oversight.

The organizing team, brings together academic and industry researchers from five European institutions, combining expertise in Federated Learning, Differential Privacy and Document Analysis and Recognition.

Team members have successfully organized previous challenges, including Document VQA (15; 20), Scene-Text VQA (1), and Text Detection, Recognition and Segmentation in images and videos (7; 6).

Dimosthenis Karatzas is an associate professor at the Universitat Autònoma de Barcelona and associate director of the Computer Vision Centre (CVC) in Barcelona, Spain, where he leads the Vision, Language and Reading research group (`http://vlr.cvc.uab.es`). He has produced more than 140 publications on computer vision, reading systems and multimodal learning. He is a prominent figure in the Reading Systems field, where he received the 2013 IAPR/ICDAR Young Investigator Award. He has received a Google Research Award (2016) and two Amazon Machine Learning Research Awards (2019, 2022). He has set up two spin-off companies to date, TruColour

Ltd, UK, in 2007 and AllRead, Spain, in 2019. Between 2018-19 he advised the Catalan government on the Catalan strategy of AI. He is a senior member of IEEE, a member of ELLIS and co-director of the ELLIS Unit Barcelona, past chair of IAPR TC11 (Reading Systems), and a member of the Artificial Intelligence Doctoral Academy (AIDA) Research and Industry Board. He created the Robust Reading Competition portal (`https://rrc.cvc.uab.es/`), established as the de-facto international benchmark in document analysis and used by more than 40,000 registered researchers.

Rubèn Tito is a Ph.D. student that joined the Computer Vision community in 2018 working on retrieval systems of natural scenes with text and handwritten text images. He later focused on Vision and Language systems, more specifically exploring the role of recognized text in the images for Visual Question Answering task in both documents and natural scenes.

Mohamed Ali Souibgui is a postdoctoral researcher at Computer Vision Center, Barcelona, Spain. He received the Ph.D. degree in 2022 from the Universitat Autònoma de Barcelona (UAB), Spain. His research focuses on document image analysis using computer vision and machine learning tools.

Khanh Nguyen is currently a PhD student in the Computer Vision Center, Barcelona, Spain. His research focuses on machine learning methods for Vision-and-Language tasks, particularly exploring the role of context and incorporate it into the image interpretation pipeline.

Raouf Kerkouche is a Postdoctoral Fellow at the CISPA Helmholtz Center for Information Security advised by Prof. Mario Fritz. His current research centers around trustworthy machine learning with a focus on private and secure collaborative machine learning as well as on private synthetic data generation. Raouf obtained his Ph.D. at INRIA, supervised by Prof. Claude Castelluccia and Prof. Pierre Genevès, where he worked on Differentially Private Federated Learning for Bandwidth and Energy Constrained Environments, with an interest in medical applications. One of his differentially private compression approaches published at UAI'21 has been included in a federated learning platform developed for drug discovery (`https://www.melloddy.eu`). He obtained his Master's degrees from Paris-Sud University and Pierre and Marie Curie University in France.

Kangsoo Jung is working as a postdoctoral researcher at the COMETE team hosted Inria. He ie working under the supervision of Catuscia Palamidessi. He received the Ph.D. degree in 2017 from Sogang University in South Korea. His research focuses on differential privacy, machine learning and game theory to address the privacy-utility tradeoff.

Marlon Tobaben is a PhD student at the Department of Computer Science, University of Helsinki, supervised by Prof Antti Honkela and affiliated with the Finnish Centre of Artificial Intelligence (FCAI), a flagship of research excellence appointed by the Academy of Finland. Marlon's research focuses on differentially private deep and federated learning.

Joonas Jälkö is a postdoctoral researcher in Professor Antti Honkela's group at the Department of Computer Science in University of Helsinki. His research focuses mainly on differential privacy applied on statistical inference and differentially private synthetic data.

Vincent Poulain d'Andecy is the head of the Yooz Research and Technologies Department since 2015. He is a graduate engineer of INSA Rennes and PhD of La Rochelle University. He started his career at ITESOFT in 1994 and has more than 25 years of experience in the development of Automatic Document Processing Systems. At Yooz, he is in charge of the AI developments with a 9-persons team, he supervises PhD and collaborative research projects in partnership with Academia like La Rochelle University and the CVC-CERCA.

Ernest Valveny received the Ph.D. degree in 1999 from the Universitat Autònoma de Barcelona (UAB), Spain. He joined the Computer Science Department at UAB in 1992 as an assistant professor and since 2002 as an Associate Professor. Since 2013 he is the Director of the Computer Science Department. He is also a researcher at the Computer Vision Center, where he is a member of the Robust Reading research unit. His main research interests are computer vision and pattern recognition, and in particular text recognition and retrieval, shape representation, document classification and graph matching. He has published more than 20 papers in international indexed journals and more than 100 papers in peer-reviewed international conferences, with more than 4800 citations. He has an h-index of 30. He has participated in a number of national and international research projects mainly related to document analysis and robust reading. He has also led several technology transfer contracts with companies, mainly related to the design and implementation of robust reading systems in open environments. He is currently a member of IAPR and of the editorial board of the International

Journal on Document Analysis and Recognition. He has served as a reviewer and member of the committee program for many of the most relevant international journals and conferences within the area of computer vision and pattern recognition (PAMI, Pattern Recognition, Pattern Recognition Letters, CVPR, ICCV, ECCV, BMVC, ...). He has actively participated in the organization of several research events.

Josep Lladós is an Associate Professor at the Computer Sciences Department of the Universitat Autònoma de Barcelona and a staff researcher of the Computer Vision Center, where he is also the director since January 2009. He is chair holder of Knowledge Transfer of the UAB Research Park and Santander Bank. He is the head of the Pattern Recognition and Document Analysis Group (2009SGR-00418). His current research fields are document analysis, structural and syntactic pattern recognition and computer vision. He has been the head of a number of Computer Vision R+D projects and published more than 200 papers in national and international conferences and journals.

Antti Honkela Professor a Data Science (Machine Learning and AI) at the Department of Computer Science, University of Helsinki. He is the coordinating professor of Research Programme in Privacy-preserving and Secure AI at the Finnish Center for Artificial Intelligence (FCAI), a flagship of research excellence appointed by the Academy of Finland, and leader of the Privacy and infrastructures WP in European Lighthouse in Secure and Safe AI (ELSA), a European network of excellence in secure and safe AI. He serves in multiple advisory positions for the Finnish government in privacy of health data. His research focuses on differentially private machine learning and statistical inference. He is an Action Editor of Journal of Machine Learning Research and Transactions on Machine Learning Research, and regularly serves as an area chair at leading machine learning conferences (NeurIPS, ICML, AISTATS).

Mario Fritz is a faculty member at the CISPA Helmholtz Center for Information Security, an honorary professor at Saarland University, and a fellow of the European Laboratory for Learning and Intelligent Systems (ELLIS). Prior, he led a research group at the Max Planck Institute for Informatics, was a postdoc at ICSI and UC Berkeley, and did his Ph.D. at TU Darmstadt. His research focuses on trustworthy AI, especially at the intersection of information security and machine learning. He is Associate Editor of the journal "IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)", regularly serves on program committees for key conferences on machine learning and S&P (e.g. ICML, NeurIPS, S&P, CCS), coordinates the "European Lighthouse on Secure and Safe AI (ELSA)" and has published over 100 scientific articles - 80 of them in top conferences and journals.

3.2 Resources provided by organizers

We will provide the benchmark site for the participants to download the dataset and submit their results. Moreover, we will also share the *"starting kit"* with the implemented baselines through a GitHub repository. Dedicated support staff for the competition will assist participants in using the baseline software and tools provided.

3.3 Support requested

N/A.

Acknowledgments

This work is supported by ELSA - European Lighthouse on Secure and Safe AI funded by the European Union under grant agreement No. 101070617. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the European Commission can be held responsible for them.

References

[1] Ali Furkan Biten, Rubèn Tito, Andres Mafla, Lluis Gomez, Marçal Rusinol, Minesh Mathew, CV Jawahar, Ernest Valveny, and Dimosthenis Karatzas. Icdar 2019 competition on scene text visual question answering. In 2019 International Conference on Document Analysis and Recognition (ICDAR), pages 1563–1570. IEEE, 2019.

[2] Chong Fu, Xuhong Zhang, Shouling Ji, Jinyin Chen, Jingzheng Wu, Shanqing Guo, Jun Zhou, Alex X Liu, and Ting Wang. Label inference attacks against vertical federated learning. In 31st USENIX Security Symposium (USENIX Security 22), Boston, MA, Aug. 2022. USENIX Association.

[3] Filippo Galli, Sayan Biswas, Kangsoo Jung, Tommaso Cucinotta, and Catuscia Palamidessi. Group privacy for personalized federated learning. In Proceedings of the 9th International Conference on Information Systems Security and Privacy. SCITEPRESS - Science and Technology Publications, 2023.

[4] Jonas Geiping, Hartmut Bauermeister, Hannah Dröge, and Michael Moeller. Inverting gradients - how easy is it to break privacy in federated learning? In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, Advances in Neural Information Processing Systems, volume 33, pages 16937–16947. Curran Associates, Inc., 2020.

[5] Robin C. Geyer, Tassilo Klein, and Moin Nabi. Differentially private federated learning: A client level perspective. CoRR, abs/1712.07557, 2017.

[6] Dimosthenis Karatzas, Lluis Gomez-Bigorda, Anguelos Nicolaou, Suman Ghosh, Andrew Bagdanov, Masakazu Iwamura, Jiri Matas, Lukas Neumann, Vijay Ramaseshan Chandrasekhar, Shijian Lu, et al. Icdar 2015 competition on robust reading. In 2015 13th international conference on document analysis and recognition (ICDAR), pages 1156–1160. IEEE, 2015.

[7] Dimosthenis Karatzas, Faisal Shafait, Seiichi Uchida, Masakazu Iwamura, Lluis Gomez i Bigorda, Sergi Robles Mestre, Joan Mas, David Fernandez Mota, Jon Almazan Almazan, and Lluis Pere De Las Heras. Icdar 2013 robust reading competition. In 2013 12th international conference on document analysis and recognition, pages 1484–1493. IEEE, 2013.

[8] Raouf Kerkouche, Gergely Ács, Claude Castelluccia, and Pierre Genevès. Constrained differentially private federated learning for low-bandwidth devices. In Cassio de Campos and Marloes H. Maathuis, editors, Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence, volume 161 of Proceedings of Machine Learning Research, pages 1756–1765. PMLR, 27–30 Jul 2021.

[9] Raouf Kerkouche, Gergely Ács, Claude Castelluccia, and Pierre Genevès. Privacy-preserving and bandwidth-efficient federated learning: An application to in-hospital mortality prediction. In Proceedings of the Conference on Health, Inference, and Learning, CHIL '21, page 25–35, New York, NY, USA, 2021. Association for Computing Machinery.

[10] Raouf Kerkouche, Gergely Ács, Claude Castelluccia, and Pierre Genevès. Compression boosts differentially private federated learning. In 2021 IEEE European Symposium on Security and Privacy (EuroS&P), pages 304–318, 2021.

[11] Bruno Koch. The e-invoicing journey 2019-2025. Preuzeto, 25:2021, 2019.

[12] Oscar Li, Jiankai Sun, Xin Yang, Weihao Gao, Hongyi Zhang, Junyuan Xie, Virginia Smith, and Chong Wang. Label leakage and protection in two-party split learning. NeurIPS 2020 Workshop on Scalability, Privacy, and Security in Federated Learning (SpicyFL), 2020.

[13] Zhuohang Li, Jiaxin Zhang, Luyang Liu, and Jian Liu. Auditing privacy defenses in federated learning via generative gradient leakage. The IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR), 2022.

[14] Virendra J. Marathe and Pallika Kanani. Subject granular differential privacy in federated learning, 2022.

[15] Minesh Mathew, Ruben Tito, Dimosthenis Karatzas, R Manmatha, and CV Jawahar. Document visual question answering challenge 2020. arXiv preprint arXiv:2008.08899, 2020.

[16] Luca Melis, Congzheng Song, Emiliano De Cristofaro, and Vitaly Shmatikov. Exploiting unintended feature leakage in collaborative learning. In 2019 IEEE symposium on security and privacy (SP), pages 691–706. IEEE, 2019.

[17] Milad Nasr, Reza Shokri, and Amir 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), pages 739–753. IEEE, 2019.

[18] Market Research. Document capture software market research report by solution, deployment, industry, region - global forecast to 2027 - cumulative impact of covid-19. `https://www.marketresearch.com/360iResearch-v4164/Document-Capture-Software-Research-Solution-32386517/`, 2022. Accessed: 2023-04-21.

[19] Štěpán Šimsa, Milan Šulc, Michal Uřičář, Yash Patel, Ahmed Hamdi, Matěj Kocián, Matyáš Skalický, Jiří Matas, Antoine Doucet, Mickaël Coustaty, et al. Docile benchmark for document information localization and extraction. arXiv preprint arXiv:2302.05658, 2023.

[20] Rubèn Tito, Dimosthenis Karatzas, and Ernest Valveny. Document collection visual question answering. In International Conference on Document Analysis and Recognition, pages 778–792. Springer, 2021.

[21] Rubèn Tito, Dimosthenis Karatzas, and Ernest Valveny. Hierarchical multimodal transformers for multi-page docvqa. arXiv preprint arXiv:2212.05935, 2022.

[22] Rubèn Tito, Minesh Mathew, CV Jawahar, Ernest Valveny, and Dimosthenis Karatzas. Icdar 2021 competition on document visual question answering. In International Conference on Document Analysis and Recognition, pages 635–649. Springer, 2021.

[23] Stacey Truex, Nathalie Baracaldo, Ali Anwar, Thomas Steinke, Heiko Ludwig, Rui Zhang, and Yi Zhou. A hybrid approach to privacy-preserving federated learning. In Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security, AISec'19, page 1–11, New York, NY, USA, 2019. Association for Computing Machinery.

[24] Aidmar Wainakh, Fabrizio Ventola, Till Müßig, Jens Keim, Carlos Garcia Cordero, Ephraim Zimmer, Tim Grube, Kristian Kersting, and Max Mühlhäuser. User-level label leakage from gradients in federated learning. Proceedings on Privacy Enhancing Technologies, 2022(2):227–244, 2022.

[25] Bo Zhao, Konda Reddy Mopuri, and Hakan Bilen. idlg: Improved deep leakage from gradients. arXiv preprint arXiv:2001.02610, 2020.

[26] Ligeng Zhu, Zhijian Liu, and Song Han. Deep leakage from gradients. Advances in Neural Information Processing Systems, 32, 2019.