Federal Register Notice 87 FR 35250, https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies, June 9, 2022

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

IBM Research

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Request for Information on Advancing Privacy-Enhancing Technologies

IBM Research

International Business Machines Corporation

July 8, 2022

July 8, 2022

Subject: Request for Information on Advancing Privacy-Enhancing Technologies.

Dear Team:

IBM Research is pleased to offer the following preliminary information in response to the Office of Science and Technology Policy's Request for Information on Advancing Privacy-Enhancing Technologies.

Our response has been prepared by:

Omri Soceanu
Director AI Security
IBM Research - Security

Nir Drucker
Principal Researcher
IBM Research - Security

John Buselli
Offering Manager
IBM Research – Security

IBM Research has a long history of working closely with United States Federal Agencies to support the development of critical security initiatives. We recognize the importance of this initiative and welcome the opportunity to continue to provide input and guidance toward this effort.

We look forward to future collaboration.

Regards,

John Buselli
Offering Manager
IBM Research
jbuselli@us.ibm.com

Request for Information on Advancing Privacy-Enhancing Technologies

IBM Research

Overview

There are numerous privacy regulations that mandate organizations abide by certain security principles when processing personal information. These principles go beyond raw data. Recent studies have shown that a malicious third party, with access to a trained Machine Learning (ML) model, even without access to the training data itself, can still reveal sensitive, personal information about the people whose data was used to train the model. It is therefore crucial to be able to recognize and protect Artificial Intelligence (AI) models that may contain personal information.

In addition to regulation mandates surrounding personal identifiable information, many organizations are enacting stronger privacy related policies and frameworks to keep their confidential and sensitive data safe. Some of the major policies and frameworks already enacted by large organizations focus on Zero Trust, Key Management, Secure Multi-Party Computation (SMPC) and Access Management approaches.

IBM Research is working on several novel techniques and tools to both assess the privacy risk of AI-based solutions, and to help them adhere to any relevant privacy requirements. We have developed tools to address the different tradeoffs between privacy, accuracy and performance of the resulting models, and for addressing the different stages in the ML lifecycle. These developments include:

AI on encrypted data - Fully Homomorphic Encryption (FHE) allows data to remain encrypted even during computation. Using FHE we are able implement a wide variety of analytics and AI solutions over encrypted data.

Differential Privacy - This method allows queries to be executed on sensitive data while preserving the privacy of individuals in the data with its robust mathematical guarantees. Differential privacy relies on random noise to protect individuals' privacy while preserving accuracy on aggregate statistics and has applications in ML and data analytics more generally.

ML anonymization - This method creates a model-based, tailored anonymization scheme to anonymize training data before using it to train an ML model, enabling to create ML models that no longer contain any personally identifiable information.

Data minimization - This technique helps to reduce the amount and granularity of features used by machine learning algorithms to perform classification or prediction, by either removal (suppression) or generalization techniques.

Privacy risk assessment - We have developed ways to assess and quantify the privacy risk of ML models and to enable comparing and choosing between different ML models based not only on accuracy but also on privacy risk criteria.

Privacy in Federated Learning - Federated Learning (FL) is an approach used in machine learning in which a group of parties (data owners), work together to train a model collaboratively without sharing training data. FL enables the exchanging and merging the parameters of locally trained models. For increased privacy this approach can be combined with other techniques such as differential privacy, homomorphic encryption and secure multi-party computation.

We offer the following for additional consideration:

Fully Homomorphic Encryption

Fully Homomorphic Encryption (FHE) is a promising solution that has been getting significant attention because of its ability to perform an evaluation of certain functions on encrypted inputs. Specifically, FHE is using different analytics such as logistic regression, neural networks (NNs), or decision trees to provide inference results on private data (i.e., customers, patients, employees, transactions). The potential of FHE to handle private data was captured e.g., by Gartner, which predicted that fifty percent (50%) of large organizations will adopt such computational models for processing in untrusted environments for multiparty data analytics by 2025 [1]. Sectors that will most likely benefit from FHE are the Health, Finance and Government sectors who can collaborate and perform complicated functions over their private data, while keeping it confidential.

Several major barriers limit the wide adoption of FHE solutions to date namely: latency, memory usage, storage, key management, regulations, and standardization. IBM Research understands these barriers and invests in fundamental and applied research for ways to solve the first three dimensions, namely, latency, memory usage and storage, while collaborating with other companies on standardizing the available FHE solutions. We elaborate next on the standardization, key management and privacy over FHE barriers.

Standardization: NIST FIPS 800-57 [2] [Section 4] ('Recommendation for Key Management Part 1 – Genera') refers to three (3) types of approved cryptographic algorithms: hash functions, symmetric-key algorithms, and asymmetric-key algorithms. Nevertheless, it does not explicitly mention FHE or refers to it. One possible reason is that until a decade ago, FHE was considered impractical, and only a small number of organizations have experimented with it. In fact, standard organizations such as NIST were asking the cryptographic community to focus on other types of cryptographic algorithms, such as light-weight cryptography [3], or post-quantum cryptography [4]. Recently, this situation has changed, where organizations such as NIST [5], The Open Industry / Government / Academic Consortium to Advance Secure Computation [6], and ISO/IEC [7] already started to consider FHE standards. In IBM we believe that having such standards and including FHE primitives as approved primitives for FIPS 140-2 [8] based solutions, may speed up the adoption of FHE solutions by many organizations.

Key Management: Cryptographic keys play an important role in cryptographic algorithms. Having unique and well-formatted keys is a prerequisite for the security guarantees that these cryptosystems provide. However, once an adversary puts its hands on these keys, the associated cryptographic scheme can no longer guarantee the confidentiality or the integrity of the key owner's data.

Generalizing the Key Management Systems (KMS) recommendations [1] to support different FHE schemes such as CKKS, TFHE, and BGV with different key characteristics involves five (5) types of keys: secret, public, evaluation, rotation, and bootstrapping keys. Each type has a unique size and usage characteristics, which a KMS solution should support. In addition, access control mechanisms to these keys should be defined based on a valid trust model and corresponding security assumptions. For example, when a hospital would like to provide a cancer detection service to its clients using some untrusted cloud environment. The hospital (data owner) should encrypt the model and then either give access to the private key to the users to use the model, or should provide some other mechanism to do so, such as multi-key FHE or proxy-re-encryption. A KMS solution should provide definitions for the hospital on how and where they are allowed to store the different keys, who should get access, and how often the keys should be rotated. While IBM is working on developing such KMS solutions, we expect wide adoption only after standards are adopted for such exchanges.

Hardware Secure Module (HSM): Current HSMs are designed to meet the requirements of symmetric and asymmetric cryptosystems that uses much smaller keys compared to FHE keys. At IBM Research, we are investigating the means to leverage typical, existing hardware for building a FHE KMS service that uses HSMs (or to suggest new HSM designs that meet FHE requirements). Standardization will likely increase the adoption of such solutions.

Hybrid encryption (transciphering): To address the issue of the ciphertext size and computational overload on edge devices, a transciphering framework, also called hybrid encryption was suggested [9]. The idea is that the IoT/edge device use some symmetric encryption scheme to encrypt its data and send the ciphertexts, together with the encryption of the key under FHE, to the server which can decrypt the data "under" FHE. This allows faster communications (at the cost of extra latency) for the decryption on the server-side. One known barrier is that current symmetric encryption schemes are not considered efficient under FHE. One possible solution is to standardize new symmetric ciphers that behave more efficiently under FHE.

AI Privacy

There is a known tension between the need to analyze personal data to drive business outcomes and the need to preserve the privacy of data subjects. Many data protection regulations, including the EU General Data Protection Regulation (GDPR) and the California Consumer Protection Act (CCPA), set out strict restrictions and obligations on the collection and processing of personal data.

Many data processing tasks nowadays involve machine learning. In recent years, several attacks have been developed that are able to infer sensitive information from trained models, including membership inference attacks, model inversion attacks and attribute inference attacks. This has led to the conclusion that machine learning models themselves should, in some cases, be considered personal information.

In 2019 the British Information Commissioner’s Office published an AI Auditing Framework which specifically mentions purpose limitation and data minimization, fairness, transparency, accountability and many more considerations. In 2020, the European

Parliament published a study on the impact of GDPR on artificial intelligence, also mentioning purpose limitation and data minimization. In 2021, UNESCO published a draft Recommendation on the Ethics of Artificial Intelligence, mentioning privacy and data protection, privacy by design and privacy impact assessments as some of the recommended practices, and is currently in the process of designing an ethical impact assessment (EIA) tool for AI. Also in 2021, the European Commission proposed a draft regulation on trust in AI, becoming the first governmental body in the world to issue a draft regulation aimed specifically at the development and use of AI. The National Institute of Standards and Technology NIST (under DoC), on a directive from the American Congress, just published an initial version of an AI risk management framework (RMF).

Recent surveys indicate that organizations are currently struggling with building AI solutions that involve personal data. In addition, security and privacy of data for ML, as well as building trustworthy and ethical AI remain great challenges. This problem is exacerbated by reports predicting that privacy-preserving techniques for AI model training will unlock up to 50% more personal data for model training and 70% more AI collaborations in industry.

We have identified three (3) main areas of research and innovation we believe will be crucial in the next few years:

1. **Privacy risk assessment of models -** this is the first step to understanding which models pose a privacy risk, enable comparing between model alternatives based on privacy criteria (and not only accuracy), and to prioritize models for further action and possible mitigation strategies. Moreover, the increased use of third party or publicly available models, as well as the emergence of AI insurance companies, increase the need for AI privacy auditing. Risk assessment can be either theoretical or empirical, however we believe that a quantitative approach is critical to enable scaling and automation of this complex and time-consuming task.

2. **Easy to consume privacy-preserving AI technology -** many privacy practices and methods require deep expertise and/or consist of invasive techniques that require significant changes to existing workflows. We believe that privacy practices must be easily incorporated into existing ML-Ops pipelines to become widely adopted. This means supplying non-invasive techniques, with standard APIs that can be added as an additional step into existing pipelines, rather than replacing or disrupting existing practices. They should also support the principle of separation of concerns, for example by applying privacy it should not require deep data science expertise, and vice versa.

3. **Compliance with data protection regulations for AI models -** as mentioned earlier, ML models are not exempt from data protection principles such as purpose limitation, data minimization and the right to be forgotten. Therefore, specially tailored solutions for applying these principles in the domain of ML must be designed and implemented. Clear guidelines on how these principles should be regarded for ML models are still lacking. If a person requests deletion of their personal information, how does this apply to ML models that were trained on this data? How can deletion be measured or verified and what

constitutes a "good enough" solution? Is exact deletion required or can approximate unlearning be applied?

For more information on this and related security topics please visit:

https://w3.ibm.com/w3publisher/ibm-research-security

****

[1] Gartner. 2021. Gartner Identifies Top Security and Risk Management Trends for 2021. Technical Report. https://www.gartner.com/en/newsroom/press-releases/2021-03-23-gartner-identifies-top-security-and-risk-management-t

[2] Elaine, B.: NIST Special Publication 800-57: Recommendation for Key Management Part 1 – General (2021). https://doi.org/10.6028/NIST.SP.800-57pt1r57

[3] NIST: LightweightCryptography https://csrc.nist.gov/projects/lightweight-cryptography (2021), last accessed 30 Sep 2021

[4] NIST: Post-Quantum Cryptography https://csrc.nist.gov/projects/post-quantum-cryptography (2021), last accessed 30 Sep 2021

[5] NIST: Toward a PEC Use-Case Suite (2021) https://csrc.nist.gov/CSRC/media/Projects/pec/documents/suite-draft1.pdf5

[6] Albrecht, M., Chase, M., Chen, H., Ding, J., Goldwasser, S., Gorbunov, S., Halevi,S., Hoffstein, J., Laine, K., Lauter, K., Lokam, S., Micciancio, D., Moody, D.,Morrison, T., Sahai, A., Vaikuntanathan, V.: Homomorphic encryption securitystandard. Tech. rep., HomomorphicEncryption.org, Toronto, Canada (November2018), https://homomorphicencryption.org/standard/2

[7] ISO/IEC: ISO/IEC 18033-6:2019 IT Security techniques — Encryption algorithms— Part 6: Homomorphic encryption (2021), https://www.iso.org/standard/67740.html10

[8] NIST: FIPS PUB 140-2: Security Requirements for Cryptographic Modules (2002). https://doi.org/10.6028/NIST.FIPS.140-212

[9] Naehrig, M., Lauter, K., Vaikuntanathan, V.: Can homomorphic encryption be practical? In: Proceedings of the 3rd ACM Workshop on Cloud Computing Security Workshop, pp. 113–124. ACM (2011)