Letter from the Special Issue Editor

The rapid advancement of Artificial Intelligence (AI) across diverse fields like healthcare, finance, transportation, and entertainment hinges heavily on data engineering/science. This discipline plays a crucial role in developing novel methods for collecting, storing, processing, and analyzing vast amounts of data, often containing sensitive personal information. However, traditional approaches such as access control and anonymization face significant challenges in handling sensitive data in the age of AI. This special issue contributes to the critical discussion on privacy-preserving technologies for data engineering/science in the age of AI and features the following seven articles.

1. *Privacy, Policy, and Compliance in Yet Another 'Age': The Necessity of Interdisciplinary Collaboration for Artificial Intelligence Applications by Bailey Kacsmar.* This article focuses on the existing domains relevant to AI governance, such as privacy, policy, and compliance, and discusses the challenges in different stages of AI and solutions that have been investigated.

2. *Data Privacy and Computation Integrity in Machine Learning Scenarios: Some Issues and Approaches by De Capitani di Vimercati et al.* While the previous paper shows the importance of AI in different fields, this paper focuses on privacy and computation integrity issues arising when data and machine learning models or tasks are shared with external parties. They develop a target-aware data anonymization technique and a solution for generating a privacy-friendly classifier that requires neither sensitive information nor information correlated with it for training a high-accuracy classifier.

3. *Beyond Data Privacy: New Privacy Risks for Large Language Models by Du et al.* The previous two papers set up the background for LLM in terms of their usages and privacy issues of ML. This paper introduces critical privacy vulnerabilities, particularly during their deployment and integration into applications. In the paper, rather than focusing on training-phase data privacy, as most existing work does, the authors explore new risks, including data leakage and malicious exfiltration enabled by LLM' autonomous capabilities. To combat these threats, the paper provides a systematic analysis of these emerging risks and calls for the development of broader, more robust defense strategies.

4. *Privacy-Preserving Federated Large Language Models: Techniques and Trade-offs by Xu et al.* Moving from the centralized LLM, this paper focuses on three specific challenges when combining Federated Learning with LLM, that are maintaining model utility amid statistical and system heterogeneity; ensuring efficiency by alleviating severe communication and computation bottlenecks; and safeguarding privacy against powerful attacks.

   The next three papers focus on specific techniques and advances for addressing various privacy challenges, including access control and differential privacy.

5. *Hyper-Scale Managed Identities and Access Control by Alagenchev et al.* This paper further discusses access control on large-scale settings and focuses on the limitations, such as scalability, adaptability, and fine-grained, context-aware control, of existing access control techniques in dynamic, distributed service-to-service architectures and presents the design and architecture of Hyper-Scale Managed Identities (HSMIs) and their integration with decentralized access control policies.

6. *Optimal Group Privacy for DP-SGD by Mahloujifar et al.* This paper addresses the long-standing challenge of tight privacy accounting for group privacy in Differentially Private Stochastic Gradient Descent (DP-SGD). While individual privacy bounds are well-established, group privacy—which

protects sets of individuals rather than just one—has historically lacked precise bounds. The researchers introduce a novel technique using "dominating pairs of distributions" to achieve tighter group privacy guarantees. They also find that sub-sampling heavily influences group privacy, meaning that two models with identical individual privacy parameters can have vastly different levels of protection for groups depending on their specific hyperparameters.

7. *Rethinking Benchmarks for Differentially Private Image Classification by Mokhtari et al.* This paper focuses on developing a benchmark to evaluate techniques for differentially private machine learning in a variety of settings, including with and without additional data, in convex settings, and on a variety of qualitatively different datasets; and creates a publicly available leaderboard for the community to track progress in differentially private machine learning.

We would like to thank all the authors for their valuable contributions. We also thank Haixun Wang for the opportunity to put together this special issue, and Jieming Shi for his help in its publication.

Shantanu Sharma[1], Xi He[2]
[1] New Jersey Institute of Technology, [2] University of Waterloo