Letter from the Editor-in-Chief

Data is considered the most valuable asset but as its volume, complexity and implications continue to grow, the tech industry must face an unprecedented range of challenges from data management to ethical obligations and regulatory pressures.

This issue of the Data Engineering Bulletin, curated by Sebastian Schelter, endeavors to address some of these challenges. For the first time, we focus on regulations shaping how data is being collected, managed, and used in the tech industry. Several papers in this issue dive into questions originating from the "right-to-be-forgotten" postulated by GDPR. A concomitant technical challenge is how to focus resources on legitimate and valuable data to maximize the business impact. For example, Davidson et al.'s work on "Disposal by Design" used e-commerce to highlight challenges and opportunities in the realm of data regulation. Applications such as e-commerce data reduction, image archiving, and relational data sampling and aggregation open the door for further research in this domain.

This issue also features an opinion piece by Ihab Ilyas and Felix Naumann, who looked into the critical question of data and model observability. For years, data quality has been a key concern and a main priority for the tech industry, but the problems have become more elusive as the industry relies more and more on machine learning models. While the situation has given rise to a new tech segment pioneered by companies such as BigEye and Monte Carlo Data, the solutions are still primitive. Ilyas and Naumann's call for action opens a new chapter of data quality and data cleaning that understands the entire data processing pipeline.

Haixun Wang
Instacart