Docket #: S26-184

Unblocking Sensitive Data with SPHERE

Advances in biomedical research and clinical medicine increasingly depend on the analysis of large, complex datasets. Artificial intelligence (AI) enables such analyses at a scale and speed that exceed human capacity, thereby accelerating scientific discovery and reshaping research across the physical and social sciences. The performance of AI systems generally improves with the volume and diversity of their training data, so access to large, varied datasets is a key limit on what they can achieve. Unfortunately, some of the most valuable data are also the most tightly protected due to privacy regulations. Furthermore, national privacy laws such as HIPAA, along with state and regional regulations that differ across jurisdictions, restrict how sensitive data may be transferred, processed, and shared. Synthetic data, which are artificial records that reproduce the statistical patterns of real data, offer a potential way to address these challenges, but current methods each involve trade-offs that sacrifice privacy, statistical fidelity, or scalability.

Dr. Zihuai He at Stanford University has developed SPHERE, a method that takes a sensitive dataset as input and produces a synthetic counterpart that preserves the original's statistical properties and supports equivalent analyses, while concealing the original records. SPHERE makes it possible to share, collaborate on, and build AI from sensitive proprietary datasets that today cannot easily leave their silos. Beyond healthcare, this novel technology has the potential to transform various sectors such as finance, government, and other privacy-constrained fields.

Stage of Development: *proof of concept*

Applications

- Analysis of large datasets

Advantages

- Across 33 real-world datasets in five scientific domains, SPHERE demonstrates superior privacy, fidelity, inference, ML utility, and speed
- SPHERE is four orders of magnitude faster than current methods used in the market
- SPHERE results in high accuracy: an AI virtual lab of four agents, run on SPHERE and on the real NHANES data reach the same scientific conclusion

Innovators

- Zihuai He

Licensing Contact

Michael Spaid

Technology Licensing Associate 2

Email