Introduction to UCLA Trustworthy AI Lab

Prof. Guang Cheng
Dept of Statistics and Data Science
Director of Trustworthy AI Lab
UCLA

CCC Talk
Dec 12, 2024

guangcheng@ucla.edu

Vison of our lab

Our lab believes that stricter regulations (such as GDPR) drive AI from 1.0 (performance-oriented) to 2.0 (trustworthiness-oriented)

Long-term collaborations with industries caring about privacy & security in DATA

Prof. Guang Cheng
Director of Trustworty AI Lab

Lab Members	Publication in AI/ML
3 Postdocs	NeurIPS, ICML
8 PhD students	ICLR, KDD, AISTATS
5 Master students	Statistics
Journals	5 Undergrad students

Major Collaborators & Sponsors

Prof. Xiaofeng Wang, Indiana Univ.,
Director of NSF Center for Distributed Confidential Computing (CDCC)

J.P.Morgan CHASE

Meta

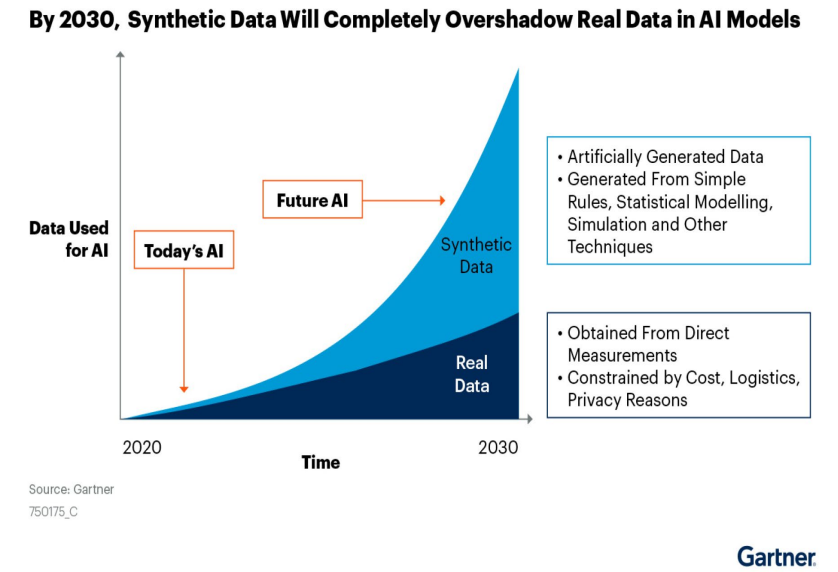
amazon

Optum Labs

CISCO

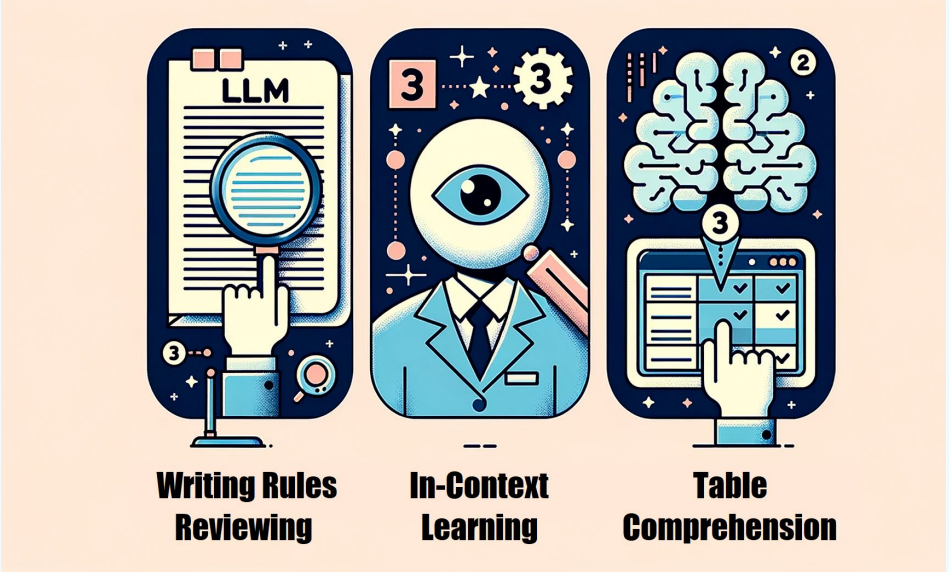
Major research areas

Data Privacy



Privacy-preserving synthetic data

Large Language Model

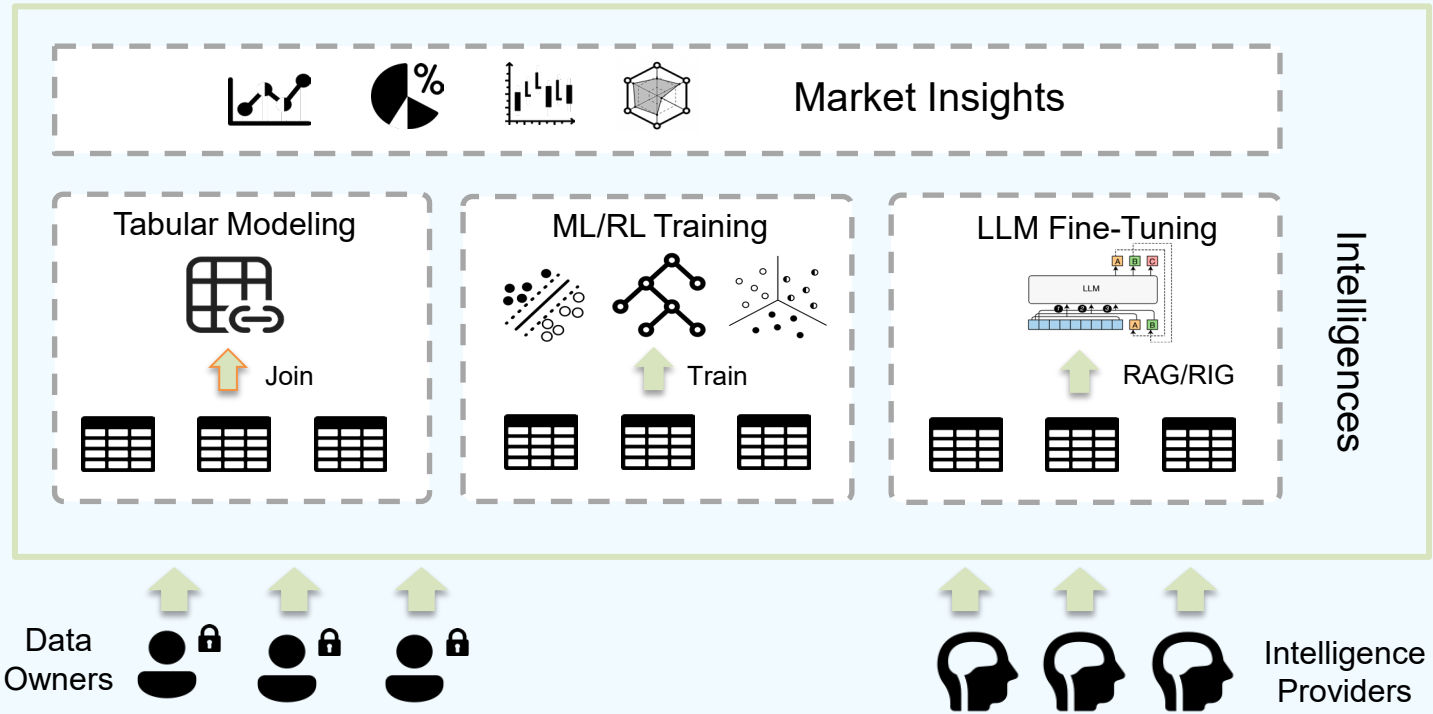


(Co-)Training of LLM in a TEE

A strong use case driving our recent research: digital marketing

- User constent: regulations such as GDPR (EU) and CCPA (California) require websites to present users with the option to accept or reject cookies upon each visit;
- Digital Markets Act: it adds further pressure on platforms enforcing stricter data-sharing practices over user data;
- Shifting towards 1P data: With 3P cookies deprecation, companies must now rely on 1P data from their own websites to collect, manage, and leverage customer insights to stay competitive.

Data collaboration platforms, e.g., data clean room, are thus developed



Future: protecting data collaboration with confidential computing

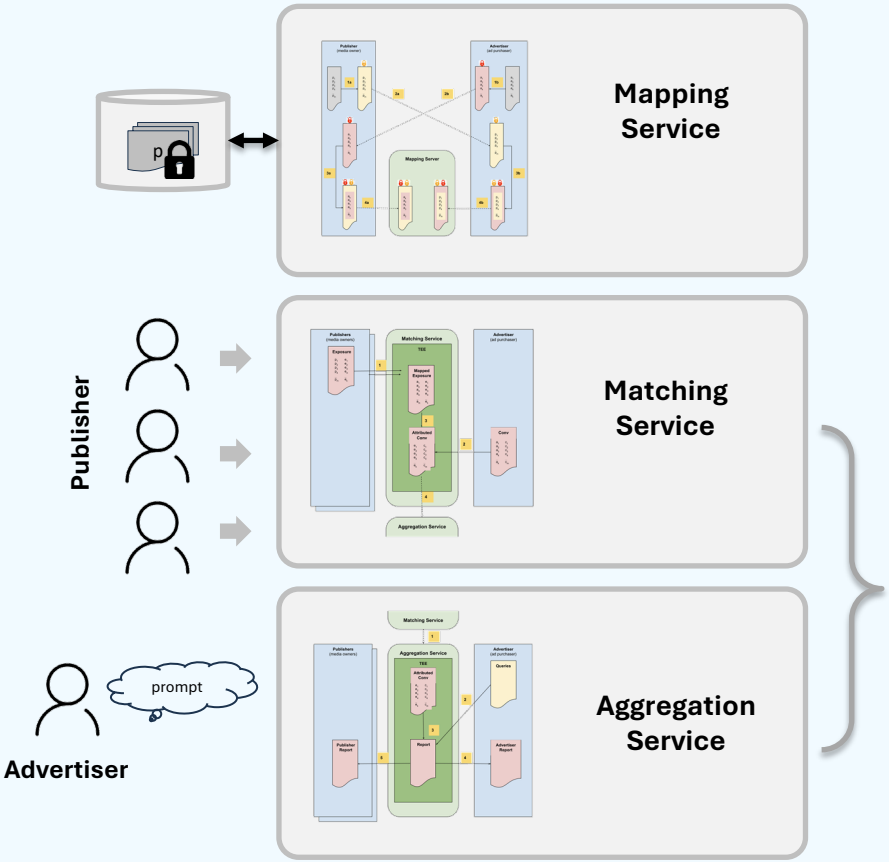
There are several components to be protected:

- Data aggregation and analytics;
- Machine/reinforcement learning model building;
- Synthetic data generation;
- Co-training of LLM

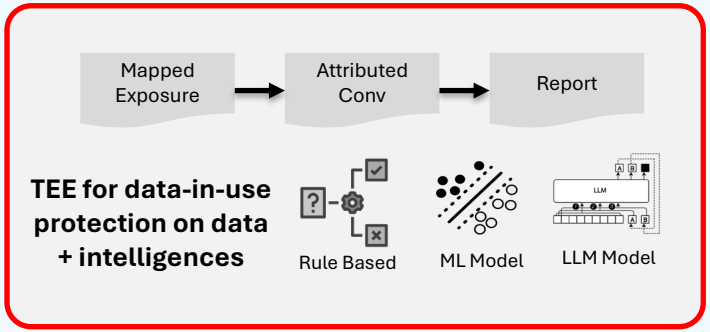
In general, we are interested in developing a TEE-based platform for secure/private data collaboration by addressing two challenges:

- Establishing a trust chain from firmware to application layer under varying policies (e.g., GDPR, HIPAA);
- Ensuring high performance and privacy across different TEEs (CPU and GPU).

TEE Provides data-in-use Protection



Use Private Set Intersection (PSI) to create ADMAP (Attribution Data Matching Protocol) IDs on blinded datasets from publisher and advertiser



Prof. Guang Cheng
Director of Trustworty AI Lab
UCLA

Contact us at
guangcheng@ucla.edu