Maaz Bin Musa

✉ maazbin-musa@uiowa.edu in maaz-bin-musa 🖥 maazbinmusa.github.io/ 📞 +1-319-821-9237

Summary

Graduating CS PhD student working in the area of online privacy and compliance measurement open to research and engineering roles.

Technical Skills

Languages | Python, C++, JavaScript, TypeScript, Bash, SQL, Terraform

Tools | RAG, Playwright, FastText, PyCharm, Docker, AI Agents, Vertex AI

Services | Azure, AWS, GCP, OpenAI, LangChain, Temporal.io, Streamlit, Huggingface, LangGraph

Experience

Privacy / Machine Learning Engineer — Aug 2024 – Dec 2024
AccuCode AI — *Iowa City, Iowa*

- Evaluated and deployed responsible **LLM automation workflows** for the purpose of extracting medical coding data as structured output from **Azure OpenAI** and **Azure custom OCR models**
- Leveraged **in-context learning**, **prompt engineering** and **fine-tuning**, to improve the performance of LLM's on specific downstream use cases
- Instrumented state of the art tools such as **Presidio**, **GLiNER** and **Microsoft Health-Data-Anonymization** to make sure the personal user data was consumed responsibly

Software Engineer — June 2024 – Sep 2024
AccuCode AI — *Iowa City, Iowa*

- Designed and deployed multiple endpoints using **FastAPI**, **Temporal.io** and **Docker containers** for in-house and production use-cases
- Conducted extensive post-processing and testing using **NER/BERT models**, **regex**, **Pytest** and **GitHub Actions** to validate extractions from LLM's
- Trained custom **text classification** models for **single and multi-page** classifications using **Azure custom models**

Projects

CCPA-SoK | *Systematic development of disclosure classification pipeline* — **Present**

- Collected privacy policies using **Playwright** and measured topic shifts using **LDA** and **LCTM** topic modeling
- Leveraged **Label Studio** and **Streamlit** to extract data and **F1/Kappa** score to measure agreement scores
- Processed policies and extracted data using **text-tiling**, **GraphSeg** and **topic-tiling** to prepare for ML pipelines
- Developed and analyzed text classification pipelines using **finetuned-BERT**, **GPT4**, **LR** and **CNN**
- Published the dataset for this project (**C3PA**) in Empirical Methods in Natural Language Processing (**EMNLP**) 2024

Forms of Disclosure | *The Path to Automated Data Privacy Audits* — **June 2023**

- **Evaluated** patterns of CCPA mandated disclosures amongst 400 data brokers using **Natural Language Processing** and **topic modeling**
- **Proposed** a standardized machine readable disclosure form which enables automated auditing
- Published in **Harvard Journal of Law and Technology** 2023

ATOM | *Ad-Network Tomography* — **July 2022**

- Designed **network-tomography** experiments using **OpenWPM**, **Ad-block filters** and **NLP toolkits** to capture personalized ad content
- Leveraged **statistical testing** and custom **LR/RF models** to uncover relationships between trackers and advertisers
- Published and presented in Privacy Enhancing Technologies Symposium (**PoPETS**) 2022
- **Runner-up for the Andreas Pfitzmann best student paper award**

CanaryTrap | *Detecting Data Misuse by Third-Party Apps on Online Social Networks* — **July 2020**

- Designed a **matrix** based scalable **honeytrap** to uncover server-side data sharing
- **Instrumented** and deployed traps on Facebook and Gmail using **Selenium**, **Python** and **Docker**
- **Caught** 16 entities **misusing** user data for **ransomware** and **personalized advertisement**
- Published in Privacy Enhancing Technologies Symposium (**PoPETS**) 2020

Education

University of Iowa — **Aug. 2019 – Jan 2025**
PhD in Online Privacy and Compliance Auditing (CGPA: 3.86) — *Iowa City, Iowa*