

MAAZ BIN MUSA

✉ maazbin-musa@uiowa.edu [in linkedin.com/in/maaz-bin-musa](https://www.linkedin.com/in/maaz-bin-musa) ☎ +1-319-821-9237

Education

University of Iowa

Aug. 2019 – June 2024

PhD in Online Privacy and Compliance Auditing (CGPA: 3.86)

Iowa City, Iowa

Experience

Privacy / Machine Learning Engineer Intern

Aug 2024 – Present

AccuCode AI

Iowa City, Iowa

- Performed **technical privacy reviews** and **DPIA's** using frameworks such as **NIST**, for technical / new features
- Evaluated and deployed responsible **LLM automation workflows** using structured output from **Azure OpenAI** and **Azure custom OCR models**
- 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 Intern

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**

Summer Intern

Jun 2016 – Aug 2016

QBatch

Lahore, Pakistan

- Designed scraping scripts using **python**, **selenium** and **beautifulsoup4** for construction websites
- Equipped the scrappers with **Javascript instrumentation** for bot/CAPTCHA detection

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 **stream-lit** 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**, **GPT4o**, **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 **dockers**
- **Caught** 16 entities **misusing** user data for **ransomware** and **personalized advertisement**
- Published in Privacy Enhancing Technologies Symposium (**PoPETS**) 2020

Technical Skills

Languages | python, javascript, typescript, bash, SQL

Tools and Services | azure document intelligence, playwright, openAI, fasttext, pycharm, dockers, temporal.io

Certifications | Data privacy fundamentals, Data science toolkit, Data privacy for information architecture