My Research Motivation

Privacy, while fundamental to freedom of expression and freedom from oppression, has become increasingly elusive. At the heart of privacy is protecting users from harm and informing them about the collection and processing of their data. The rapid advancement of AI and large language models (LLMs) and their integration into everyday technologies have significantly reshaped the privacy landscape. As these models are increasingly used to automate more tasks, they process vast amounts of diverse data, making sensitive inferences about users and blurring the boundaries between data usage purposes. This expansion not only complicates how personal information is collected and utilized but also poses new challenges for individual privacy and autonomy. Motivated by these concerns, my research aims to explore how LLMs are being utilized for corporate and government surveillance and to develop strategies to protect people's privacy in this rapidly changing environment. My primary research objectives are to (1) develop frameworks that facilitate the exploration of the various surveillance threats posed by AI, and (2) create systems that use AI to protect user data privacy.

Brief Summary of Past Projects

My earlier work focused on enhancing correctness and transparency in privacy practices. I developed systems to automatically detect violations in data privacy management, such as website cookie settings, browser extensions, and third-party tracker opt-out mechanisms. For instance, our analysis of 47.2k Chrome Web Store extensions uncovered 820 with misleading privacy disclosures [8]. Similarly, we found that 11 out of 2.9k online trackers did not comply with their own opt-out policies [11], and cookie consent mechanisms often mishandled user consent, especially outside GDPR regions [5]. Most recently, I conducted a measurement and compliance study with the Android Automotive operating system, where we found that car manufacturers were collecting sensitive vehicle data [4]. These findings underscore the urgent need for more transparent and reliable privacy practices.

To address privacy concerns in various contexts, I developed user-centric tools like Face-Off [12], a framework that introduces perturbations to faces in images to counteract recognition systems. My exploration into fairness and security in face recognition models led to methods for analyzing and mitigating biases and vulnerabilities, such as disparities in anti-face recognition systems across demographics [9], and enhancing adversarial robustness in deep neural networks for cyber-physical systems [13]. I also developed Confidant [10], a privacy controller for social robots that manages conversational privacy using NLP models based on contextual cues like sentiment and topic. Another system, Steward [6], uses LLMs to interact with and navigate websites, performing user interface actions, with speed and cost-efficiency. This system was developed with the intention of automatically reporting content, and auditing recommendation algorithms and online advertising.

These projects have laid the groundwork for my current focus on the privacy challenges posed by LLMs and AI-driven surveillance.

Ongoing and Future Work: Surveillance by LLMs

LLMs, especially multi-modal ones, can process vast quantities of diverse data types, enabling them to draw sensitive inferences about users and make decisions in various capacities. Depending on the extent of data processing, the use of LLMs for user profiling and surveillance is at risk of infringing upon indivudals' autonomy. The central question of my thesis is: *How will LLMs be used for corporate and government surveillance, and what can we do to protect people's privacy?*

LLM Surveillance Capitalism: Chatbot Advertising and User Profiling

Companies are embedding LLM-generated advertisements into chatbots, raising ethical concerns that are not yet fully understood. In our study [2], I examined how users perceive personalized ads generated by

these models.
I developed a chatbot advertising system and conducted experiments with users under three conditions: control, ads, and disclosed ads, using GPT-4o and GPT-3.5 models. Our system mimicked targeted advertising by incorporating simulated bidding, topic matching, user profiling, and demographics. Our evaluation revealed that chatbots effectively placed products and brands, encouraging user engagement subtly. Products in the advertising conditions were 19.05% more likely to be positively perceived. Users who knew that these product placements were ads generally viewed the chatbot as more biased, irrelevant, intrusive, and less trustworthy. Our user study also revealed that advertising disclosure links were insufficient privacy controls. Many more participants attempted to control ad settings by requesting the chatbot to stop or asking questions about the ads.
I am also concluding a project where ChatGPT users upload their conversation history to our website, which processes the data to construct detailed user profiles with specific inferences [3]. This highlights LLMs' ability to generate accurate inferences from conversational data, raising significant privacy concerns.

LLM Shoulder Surveillance: Automated Shoulder Surfing Attacks

Smartphones are widely used in public, and users wish to keep on-screen information private. With the rise of smart glasses, a new threat emerges: automated shoulder surfing attacks using LLMs and computer vision [1].
Our research proposes to create an attack that allows smart glasses to read and interpret nearby smartphone screens without the smartphone user's knowledge. The project leverages screen detection and upscaling, automated PIN stealing, and content snooping. These approaches will use machine learning techniques such as vision transformers and OCR to extract sensitive information like PINs, texts, photos, and other personal data.
To counter this impending problem, I designed Eye-Shield [7], a patented software solution that manipulates on-screen content to appear clear up close but blurred or pixelated from a distance or wider angles, reducing the risk of such shoulder surfing attacks.

Real-Time Vision LLM Surveillance

Advancements in vision-language models have amplified surveillance capabilities, raising significant privacy concerns. This research idea explores how various camera platforms—such as smartphones, smart glasses, surveillance systems, and autonomous vehicles—can facilitate real-time information gathering about individuals. I will develop cross-platform methods to collect and analyze visual data using object detection, tracking, geolocation inference, and facial recognition. By integrating these techniques into a comprehensive surveillance system with an LLM-enabled interface and visualization tools, I aim to expose potential risks and aid in developing effective countermeasures.

LLM-Powered Right to Access: Visualizing Data Privacy Requests

To empower users to control their personal data, I will develop a tool that identifies which entities hold their data and automates data deletion requests. Leveraging language models, I aim to create a personalized privacy assistant that adapts privacy settings, provides tailored privacy suggestions, facilitates data rights requests, and visualizes data flows. This tool seeks to make data privacy more accessible, enabling users to manage their personal information effectively.

Future Directions

Moving forward, my research will examine integrating LLM agents into cyber-physical systems like robotics, autonomous vehicles, smart glasses, and IoT-enabled buildings. For example, buying a car or smart glasses should not implicitly allow companies to analyze camera data via vision transformers to deduce behaviors or interests. As AI systems evolve rapidly—often outpacing privacy regulations and protective technologies—my goal is to prevent misuse by developing scalable protection and opt-out mechanisms for cyber-physical systems, ensuring individuals maintain control over their data and privacy.

References

[1] **Brian Tang** and K. G. Shin, "Shoulder surveillance: Ai-automated shoulder surfing attacks with smart glasses," 2025.

[2] **Brian Tang**, K. Sun, N. T. Curran, F. Schaub, and K. G. Shin, "It lied to me: Implications of injecting personalized advertising into large language model chatbots," in *Under Review: ACM CHI Conference on Human Factors in Computing Systems*, 2025. [Online]. Available: https://arxiv.org/abs/2409.15436.

[3] **Brian Tang**, Q. Zhu, and K. G. Shin, "You are what you prompt: Privacy risks in conversations with chatgpt," 2025.

[4] B. Gozubuyuk, **Brian Tang**, M. D. Pesé, and K. G. Shin, "I know what you did (in your car) last summer: Privacy implications of android automotive os," in *Under Review: 25th Privacy Enhancing Technologies Symposium*, 2025. [Online]. Available: https://arxiv.org/abs/2409.15561.

[5] **Brian Tang**, D. Bui, and K. G. Shin, "Navigating cookie compliance across the globe," in *Under Revision: 25th Privacy Enhancing Technologies Symposium*, 2024.

[6] **Brian Tang** and K. G. Shin, "Steward: Natural language web automation," 2024. [Online]. Available: https://arxiv.org/abs/2409.15441.

[7] **Brian Tang** and K. G. Shin, "Eye-shield: Real-time protection of mobile device screen information from shoulder surfing," in *32nd USENIX Security Symposium*, 2023. [Online]. Available: https://rtcl.eecs.umich.edu/rtclweb/assets/publications/2023/usenix23-tang.pdf.

[8] D. Bui, **Brian Tang**, and K. G. Shin, "Detection of inconsistencies in privacy practices of browser extensions," in *44th IEEE Symposium on Security and Privacy*, 2023. [Online]. Available: https://www.bjaytang.com/pdfs/ExtPrivA.pdf.

[9] H. Rosenberg, **Brian Tang**, K. Fawaz, and S. Jha, "Fairness properties of face recognition and obfuscation systems," in *32nd USENIX Security Symposium*, 2023. [Online]. Available: https://arxiv.org/abs/2108.02707.

[10] **Brian Tang**, D. Sullivan, B. Cagiltay, V. Chandrasekaran, K. Fawaz, and B. Mutlu, "Confidant: A privacy controller for social robots," in *17th ACM/IEEE International Conference on Human-Robot Interaction*, 2022. [Online]. Available: https://arxiv.org/abs/2201.02712.

[11] D. Bui, **Brian Tang**, and K. G. Shin, "Do opt-outs really opt me out," in *29th ACM Conference on Computer and Communications Security*, 2022. [Online]. Available: https://dl.acm.org/doi/10.1145/3548606.3560574.

[12] V. Chandrasekaran, C. Gao, **Brian Tang**, K. Fawaz, S. Jha, and S. Banerjee, "Face-off: Adversarial face obfuscation," in *21st Privacy Enhancing Technologies Symposium*, 2021. [Online]. Available: https://arxiv.org/abs/2003.08861.

[13] V. Chandrasekaran, **Brian Tang**, N. Papernot, K. Fawaz, S. Jha, and X. Wu, "Rearchitecting classification frameworks for increased robustness," 2020. [Online]. Available: https://arxiv.org/abs/1905.10900.