Motivation

Privacy, while fundamental to freedom of expression and freedom from oppression, has become increasingly elusive. Emerging artificial intelligence (AI) technologies have cast a shadow over the future of privacy. At the heart of privacy is protecting users from harm and informing them about the collection and processing of their data. Users should know if and when a service/website misuses their data and be able to protect themselves. Legal measures like the GDPR and CCPA have been instituted for privacy protection and AI regulation, yet regulators face the task of keeping pace with rapid technological advancements. With AI evolving rapidly, service providers bear the brunt of the responsibility to self-regulate and uphold user privacy. The verification of privacy practices' correctness, consistency, and legality requires extensive debugging, network traffic analysis, and consultation with privacy experts.

Therefore, my objectives are to (1) develop a framework that facilitates the exploration of various defenses against AI data misuse, and (2) create software systems that use AI to protect user data privacy.

Past Projects

In previous work, I sought to address the issues of correctness and transparency in privacy policies, by creating systems to automatically detect violations in data privacy management systems such as website cookie settings, browser extensions, and third-party tracker opt-out mechanisms. In our comprehensive analysis encompassing 47.2k Chrome Web Store extensions, we uncovered a pervasive problem of misleading privacy disclosures, with 820 extensions found to contradict their own privacy statements [1]. Similarly, our investigation into online trackers revealed that 11 out of 2.9k trackers did not comply with their stated opt-out policies, underscoring a profound trust issue in online tracking practices [2]. Furthermore, analysis of cookie consent mechanisms exposed significant inconsistencies in how user consent is managed and respected, particularly noting that regions outside the GDPR exhibit a higher rate of consent violations [3]. These findings stress the urgent need for more transparent and reliable privacy practices, and methods for mitigating the discrepancies between stated policies and actual practices.

In my research, I have developed several user-centric privacy tools aimed at enhancing user privacy in various contexts. These include Face-Off [4], a privacy-preserving framework that introduces strategic perturbations to faces in images to counteract face recognition systems. To protect sensitive information on mobile devices from "shoulder surfing" attacks (peeking or recording screens over the victim's shoulder), I designed Eye-Shield [5], a novel patented software-based solution that manipulates on-screen content to appear clear and legible at close distances while becoming blurred or pixelated from further away or at wider angles. I also proposed a privacy controller for social robots, CONFIDANT, that employs natural language processing models to understand and manage the privacy of information shared in conversational interactions, based on contextual cues such as sentiment, relationship, and topic. Lastly, my exploration into the fairness and security implications of machine learning systems has led to the development of methods to analyze and mitigate biases and vulnerabilities in these systems. This includes investigating the performance of anti-face recognition systems across different demographics [6], revealing disparities in the effectiveness of privacy-preserving techniques. Additionally, my work on enhancing the adversarial robustness of deep neural networks without compromising accuracy proposes hierarchical classification schemes that leverage invariant features in cyber-physical systems to improve system security, such as in safety-critical applications like autonomous vehicles [7].

Ongoing and Future Projects

The introduction of large language models (LLMs) has complicated the privacy landscape by expanding the ways in which data can be used, blurring the boundaries between data usage purposes. These models,

typically multi-modal, can process a broad quantity and variety of data types. Thus, they can draw sensitive inferences about users from diverse data sources, and make decisions in various capacities. For instance, when a user consents to data collection for marketing and online behavioral advertising, it should not implicitly authorize service providers to use LLMs to deduce their psychological or emotional states, nor their sensitive traits, for advertising purposes.

In response to these concerns, my research delves into the risks of employing LLMs in advertising and data processing. I am focusing on how users perceive personalized advertisements and the profiles these models generate in a study on embedding advertising into chatbots [8]. Concurrently, I am creating an automated framework for interacting with websites via LLMs. This system, called Steward [9], can model website contexts, perform user interface actions, and monitor context-specific data collection with state-of-the-art accuracy, reliability, and speed. It has the potential to act as a universal API for web services and enable future research on recommendation algorithms, data privacy, and online advertising.

Looking forward, my future projects will examine the integration of LLM agents within cyber-physical systems like autonomous vehicles, smart glasses, and IoT-enabled buildings. For example, the act of purchasing cars or smart glasses should not implicitly permit companies to analyze one's camera data through GPT-4's vision transformer to deduce behaviors or interests. Irrespective of these concerns, AI systems are rapidly evolving, outpacing privacy regulations and the development of privacy-enhancing technologies. My goal is to prevent such misuse by creating scalable protection and opt-out mechanisms tailored for cyber-physical systems.

References

[1] 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.

[2] ——, "Do opt-outs really opt me out," in *29th ACM Conference on Computer and Communications Security 2022*, 2022.

[3] **Brian Tang**, D. Bui, and K. G. Shin, "Detection and analysis of cookie violations," in *In Review: 30th ACM Conference on Computer and Communications Security*, 2023.

[4] 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`.

[5] **Brian Tang** and K. G. Shin, "Real-time protection of mobile device screen information from shoulder surfing," in *Major Revision: 32nd USENIX Security Symposium*, 2023.

[6] 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`.

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

[8] **Brian Tang**, N. T. Curran, F. Schaub, and K. G. Shin, "Embedding advertising in llm chatbots: Risks and ethical considerations," in *In Preparation: ACM CHI Conference on Human Factors in Computing Systems*, 2025.

[9] **Brian Tang** and K. G. Shin, "Steward: Natural language web automation," in *Under Submission: The 30th Symposium on Operating Systems Principles*, 2024.