My research interests are at the intersection of data privacy and machine learning (ML), focusing on federated learning (FL) and differential privacy (DP). I've always wanted to study this area with greater depth and pursue a career as a professor, and I've been working hard towards this career, via research, teaching, and service to the academic community. I'm applying for the EECS PhD program at MIT as it presents an unparalleled opportunity to collaborate with experts, further my research, and lay the groundwork for a fulfilling career in academia.

Research in federated learning. My journey into data privacy and ML began with a research internship in summer 2021 at the Institute for Mathematical Science (IMS). During the project, I led a team designing a vertical FL system for dual-party financial settings, to meet the requirements set by our sponsor, Google. Our goal was to enable organizations like banks and fintech companies to collaboratively train models on overlapping users without compromising privacy. To this end, I applied homomorphic encryption to our FL system, ensuring all exchanged data remained encrypted, and calculations were performed on ciphertext. This experience solidified my interest in FL, and I continued exploring this field as a research assistant hosted by Prof Xiaokui Xiao. Here, in my junior year, I investigated the privacy of split learning (SL), an emerging alternative to FL. My project critiqued previous attacks on SL, revealing their reliance on unrealistic assumptions and vulnerabilities in simpler models. I developed practical attacks that yielded significantly better results under realistic assumptions. This work earned me the Outstanding Computing Project Prize from NUS School of Computing, and led to a first-authored paper, currently under review for WWW '24 with generally positive reviews and grades.

Research in differential privacy. After my junior year, I found my research interests leaning towards theoretical problems and in particular, DP as a rigorous framework to quantify information leakage. Supervised by Professors Xiaokui Xiao and Vincent Tan, I completed my honors thesis on graph neural networks (GNNs) with link local privacy. We aimed to safeguard privacy in GNN training via local DP, where each node/user perturbs its data with noise before transmission to the server, ensuring that the server receives only differentially private information. To address the inherent privacy-utility trade-off, I innovatively proposed to use Bayesian estimation to estimate the graph topology, striking a privacy-utility balance with theoretical guarantees. This work has a broader potential impact, including privacy-preserving contact tracing systems. An abstract of this work was published and presented in SIGMOD '23, where I was awarded first-place in the SIGMOD Student Research Competition (SRC), and qualified to compete in the ACM SRC grand finals. Additionally, a full research paper first authored by me was published and presented in CCS '23. This work has also earned me the University-level Outstanding Undergraduate Researcher Prize, recognizing exceptional undergraduate research contributions.

In the meantime, I'm curious about the limitations of DP as a notion to quantify information leakage on individual data points. Driven by such curiosity, we studied the data distribution leakage in differentially private FL. We found that, despite DP guarantees, it's still possible to infer the data distribution of individual organizations, a concern in scenarios where such information could divulge sensitive business insights. This exploration opens new avenues in data privacy and underscores the need for stronger protection. The paper, co-authored by me, is currently undergoing revision at TKDE.

Upon graduating from NUS with double honors degrees in CS and math, both with highest distinction, I continued as a full-time RA. While I continued working on FL and DP, my focus expanded to include ML theory. Currently, I'm working on kernelized bandits, trying to develop Gaussian process optimization algorithms in a federated setting, with the aim of theoretically bounding the cumulative regret and communication cost.

Future work. As a graduate student, I'd like to improve privacy-preserving ML and explore a wide range of privacy notions. To this end, I am eager to work with Prof Srini Devadas and am particularly interested in his recent work on improvements to DP-SGD via novel optimizer designs that reduces clipping bias (S&P 2023) and adding anisotropic noise across principal subspaces (CCS 2023), as well as PAC privacy (CRYPTO 2023), a new privacy notion which allows automatic privacy measure and control. I've always believed that DP is a somewhat pessimistic protocol that focuses on worst-case but potentially rare inputs. PAC privacy addresses this by taking the prior distribution that generates the data points into account. Although my background is more aligned with algorithmic (differential) privacy, I'm also interested in exploring the cryptographic side of privacy, such as homomorphic encryption, multi-party computation and private information retrieval, which Prof Devadas has also extensively worked on. I have exchanged some early thoughts on related topics with Prof Devadas, and I've had the pleasure of talking to his students, Simon and Kevin, at CCS to learn about the group. Additionally, I'm also passionate theoretically study the fundamental privacy-utility trade-off in ML, and I'm interested Prof Martin Wainwright's work on information-theoretical bounds of local DP, and his recent work on federated optimization (NeurIPS 2020) and adversarial attacks (S&P 2020). Finally, data privacy is a field deeply rooted in the database community, and privacy can only be achieved via careful design and implementation of database systems. Hence, I am also interested in Prof Samuel Madden about his recent project, ATLANTIC (VLDB 2021), on building efficient and accurate database systems with differential privacy guarantees.

Looking ahead, I am eager to continue my research as a graduate student. The presence of distinguished experts in the field and a culture that supports research make MIT an ideal setting for my academic pursuits. Additionally, the high regard with which my advisor and mentor, Prof Vincent Tan – himself an esteemed MIT alumnus – speaks of the Institute, further fuels my aspiration to embark on my PhD journey here. I'm confident that my academic foundation and passion for research will enable me to make meaningful contributions and thrive in the challenging yet stimulating environment at MIT.