

Why you are interested in applying for the Gates Cambridge Scholarship and how you meet the four main criteria

I was always interested in stories. From historical anecdotes to literary tales, I've found that stories can humanize ideas and elicit empathy, making distant ideas seem near.

My interest in computer science (CS) came later and was rooted in my fascination with how algorithms model complex tasks. Although interesting, work in CS felt removed from reality, as computer programs seemed distant from humans. My perception changed after reading stories about the impacts of AI on real-world issues, both positive and negative. On one hand are applications like the kidney exchange program at Facebook, which uses big data to deliver life-changing results, and on the other are racial bias issues surrounding the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) recidivism algorithm. These stories convinced me of the momentous impacts AI has on real people and real social problems, and drove me to focus on researching approaches to AI that account for social goods. This led to projects on a variety of topics, including ride-pooling fairness, human-AI collaboration, and toxicity detection in online communities.

While I was interested in AI for social good, I was initially unsure whether to approach this as a researcher or an engineer. I experimented with research during my time working on online toxicity at Carnegie Mellon. I aimed for a publication, and after submitting an initial, highly technical draft to my mentors, they instructed me instead to tell the story of the research project in the paper. Puzzled, I edited my draft, restructuring it in a narrative style. Realizing that writing about research is a form of story-telling enabled me to fall in love with this humanistic and intellectually liberating craft.

If research is the process of finding and telling new stories, teaching is the process of relating discoveries to students in personally meaningful ways. I was drawn to teaching because of its ability to enact social change; achieving social good requires both research and broad, community-oriented outreach. For my first job, I helped local middle and high school students run a cybersecurity club, teaching them the fundamentals of Linux and scripting. I later taught reading to elementary schoolers, math to middle schoolers, and led classes on programming languages and coding interviews at my university. For all of these classes, I'm able to teach material by telling stories students relate to, such as anecdotes from my own interview experiences. My love of stories led to a passion for research and teaching, and because I hope to inspire others similarly, I intend to become a professor.

Pursuing advanced CS research at Cambridge will enable me to work with Dr. Jon Crowcroft on privacy-related issues in data markets, to work with world-class researchers, to take advanced classes that complement my research, and to collaborate with other Gates-Cambridge scholars.

Provide up to 5 academic awards

1. Goldwater Scholarship
2. Philip Merrill Presidential Scholar
3. AAAI Undergraduate Consortium
4. Iribe Scholarship
5. Presidents Scholarship

Select two referees

1. Bogdan Vasilescu
2. John Dickerson

Select up to 5 employment histories

1. MIT Lincoln Lab
2. Facebook
3. Carnegie Mellon
4. UMD Research
5. UMD Teaching Assistant

Career Goals (1000 characters)

As a professor, I aspire to study high public-impact applications of machine learning (ML), with an emphasis on their effects on social goods. Examples of applications include using ML for better clinical prognoses across racial groups and in criminal justice. To tackle these issues, I plan to improve the robustness and interoperability of deep learning models, then probe models to understand why fairness issues arise. During my MPhil, I plan to study techniques to mitigate privacy issues associated with artificial intelligence (AI) by developing data markets, while also learning about the equitable use of AI. My long-term goal is to produce impactful research as a professor, and to this end, I plan to work with organizations to assist with their use of ML and ensure its equitable use. Examples include working with the World Wildlife Federation to develop ML algorithms that position security guards to minimize poaching or working with Safe Place for Youth to develop interventions.

Additional information to support application (200 words)

Covid 19 has restricted my ability to travel and attend conferences, limiting my ability to network. I was scheduled to give presentations at the ICSE 2020, IJCAI 2021, and NeurIPS 2020

conferences, though because of the pandemic, these conferences transitioned online, which made attending talks and meeting researchers more difficult.

Luckily, my research and internships have not been significantly impacted by the pandemic. My research groups transitioned online, and my internship at Facebook transitioned to work from home. While the pandemic has limited my ability to meet other researchers through conferences, it has not significantly affected my research or internship plans. Additionally, because of the switch to an online environment, I was able to attend research talks online about natural language processing and artificial intelligence through my university's research talks program.

CV

Have you made any additional applications to the University of Cambridge or any other institutions?

1. MIT
2. Stanford
3. California Berkeley
4. Harvard
5. Carnegie Mellon

Why have you decided to apply for the M.Phil in Advanced Computer Science, and what interests you in particular? (1500 characters)

An MPhil in Computer Science from Cambridge allows me to further understand artificial intelligence (AI), machine learning (ML), and their impacts on social goods.

I'm interested in courses that teach me more about AI and ML, such as advanced robotics and probabilistic machine learning. Advanced robotics with Amanda Porok covers human-robot interaction and robot learning, and excites me about the possibility of human-AI collaboration. Probabilistic machine learning gives me a mathematical foundation, and probabilistic proofs are a powerful tool for proving theorems. Both of these classes complement my research, as they give me a theoretical foundation necessary for understanding prior literature.

I'm also excited to work on research with people such as Jon Crowcroft and Mateja Jamnik. I'm interested in Jon Crowcroft's research on data markets as a mechanism to combat privacy issues, and I would like to extend this work so data markets cover heterogeneous preferences for different data types. Additionally, Mateja Jamnik's research uses AI and ML to understand environmental trends, connecting with my desire to use AI for social good. My research into human-AI collaboration and

fairness has given me practice designing ML and AI algorithms, preparing me for Cambridge. Obtaining an MPhil in Computer Science from Cambridge allows me to take interesting courses, pursue research opportunities, and will prepare me to become a professor.

Please describe briefly your programming experience: (1000 characters)

I gained experience with a variety of programming languages through my courses, research, and internships. In school, I learned C, Ruby, Python, and Ocaml, using libraries including PyTorch and Scikit-learn. I learned these through courses such as computer vision, programming languages, and databases. I am also a teaching assistant for a class on programming languages, assisting students with understanding functional programming and designing language interpreters. For research, I use web development languages including React to design user interfaces, and I use Python to design machine learning and deep learning algorithms. At Facebook, I worked on developing a web UI using React, SQL, and Hack (a PHP variant) to debug feed ranking systems. For this project, I went through rounds of UI iteration and ended up writing many 1000s of lines of code. At Lincoln Labs, I used Jupyter notebooks to develop deep learning models in PyTorch, using the Pillow library to parse images.

Grade achieved in A-Level mathematics, International Baccalaureate, or international high school equivalent: (300 characters) (Highest level of math training)

Received a 7 on HL math, and in college, took real analysis, partial differential equations, probability theory, and statistical methods. Currently taking a class on numerical methods, and have taken computer science classes on deep learning, mechanism design, quantum computation, etc.

Please write a brief research project proposal of at most 3000 characters:

User data is critical for machine learning (ML) and artificial intelligence (AI) applications such as targeted advertising, leading to the creation of data markets where user data is bought and sold. These markets largely benefit data trading corporations that own user data, as consumers lack information or control over how their data is used and about data sales. While some uses benefit consumers - for example, data sold to medical companies help develop treatments - consumers lack adequate compensation and control over their data.

Consumers need privacy-preserving data markets where individuals are compensated for data sales and privacy preferences are respected. However, real-life privacy preference data is scarce and price dependencies between consumers, where the sale of one person's data can lower the value of others' data, makes calculating optimal prices infeasible. While difficult, developing privacy-preserving data markets allows users to take control of their data.

Professor Jon Crowcroft's research on privacy preservation aims to create privacy-preserving data markets. He analyzed the effects of network externalities, which arise from correlations between users, and concluded that these externalities result in a "race to the bottom" as user data valuations go to zero, hindering the development of privacy-preserving markets (Pal et al.). To counter this, Professor Crowcroft determines prices while maintaining group privacy (Pal et al.).

Two areas for future exploration are realistic privacy preferences and data heterogeneity. In "Preference-Based Privacy Markets", Professor Crowcroft asked whether we can collect realistic human privacy preferences to supplant synthetic data. I propose collecting user preferences for privacy compensation, which identifies what users think is adequate compensation for privacy loss, for different data types. For example, users could value location data privacy more than health data privacy, which could lead to markets restricting location data sales.

Incorporating heterogeneous privacy valuations for different data types requires both a compensation vector of valuations and a matrix representing correlations between data types. These correlations indicate that the sale of one type of data changes the value of another type. One method to determine data prices would be to price each data type independently, but this ignores correlations between data types that could lead to market inefficiencies. For example, users might want to increase the value of location data if health data is sold. To counter this, we can iteratively develop approximately optimal prices using reinforcement learning techniques like Markov decision processes (MDPs).

I can build on the material I learned from my class on mechanism design, using concepts such as Stackelberg games to construct markets. By working with Professor Crowcroft, I can learn more about data markets, and combat privacy-related issues.

Module Choices

Michaelmas Term

1. Natural Language Processing
 - a. No pre-requisites, but have taken graduate computational linguistics
2. Large-scale data processing and optimization
 - a. No pre-requisites
3. Probabilistic Machine Learning
 - a. Have taken Probability Theory, and dealt with probability in graduate deep learning course, graduate quantum computing course

Lent Term

1. Advanced Robotics
 - i. No Pre-requisites

2. Advanced topics in machine learning
 - i. Have taken graduate deep learning, and will have taken probabilistic machine learning by that point

Specialization

1. Machine Learning & Artificial Intelligence
2. Human-centered computing