Short Questions (~300 words)

How did you choose your field and what are the primary expectations of your future career?

I spent the first part of college jumping between various machine learning (ML) and artificial intelligence (AI) projects, realizing my research direction for graduate school through a project on fairness in rideshare with Prof. John Dickerson. For the project, I improved the fairness of rideshare matching algorithms, which suffer from driver-side income inequality and rider-side discrimination, through new fairness-based objective functions. Performing this research taught me about fairness-related issues in AI, ways to mitigate them, and the potential of AI for social good, inspiring me to further explore ways AI can help humanity by becoming a professor.

Applications of AI that interest me include healthcare and social networks. I was inspired to pursue these applications after attending talks that detailed the use of time series algorithms to predict patient outcomes and the use of vision algorithms to detect ailments using magnetic resonance images (MRIs). The talks showed me how AI could impact real-world health outcomes, motivating me to study this in the future. I aim to combine future research in healthcare with my experience in fairness; algorithms for healthcare need to be fair across demographic lines so all groups can be serviced fairly. To ensure that the algorithms I develop are employed in practice, I plan to work with local hospitals to ensure AI is being used effectively and equitably.

Outside of research, I intend on teaching and mentoring. I plan on doing this through both outreach programs to local schools and university-based teaching. Ensuring AI is used for social good requires not only research insight but also education efforts to ensure that diverse groups of populations have access to and can understand these technologies. To address this, I plan on working with local schools to increase access to computer science education, especially for students from underrepresented groups. Additionally, I plan to teach classes at a university. These classes typically face a steep learning curve, causing many students to drop from computer science, so I plan to provide resources to help struggling students.

How does your proposed field of study and career constitute an application of the physical sciences or engineering?

AI research involves the development and analysis of complex models, which requires solving engineering challenges. For example, when I worked at the MIT Lincoln Labs during the summer, I had to engineer new algorithms to improve human-AI collaboration through fine-tuning. To do this, I first analyzed code to incorporate fine-tuning into learning-to-defer algorithms, which partition tasks between AI and humans. This is tricky as learning-to-defer algorithms are sensitive to hyperparameters, and data is relatively sparse. I addressed the former challenge through hyperparameter searches, and the latter with semi-supervised learning. To test these methods, I developed a synthetic dataset, which used learning-to-defer algorithms to determine when humans should retake control over autonomous vehicles. Through my research, I had to solve a set of engineering challenges, many of which involved configuring models to new scenarios.

Engineering challenges in AI are not limited to research, as real-world AI development requires solving engineering problems. When deploying AI to real-world scenarios, such as healthcare or medical settings, there are practical problems that need to be taken care of through engineering solutions. First, data needs to be modified so it can be easily used by removing null fields and transforming categorical data through data engineering. Additionally, training AI algorithms is a computationally-intensive process, requiring many graphical processing units (GPUs) and specialized servers. In the medical domain, clients such as hospitals lack ownership of these resources, and instead deploy and run training scripts to cloud services such as Amazon Web Services. Doing so is challenging, as it involves the creation of specialized bash scripts and monitoring programs. Additionally, determining how to structure algorithms for AI requires trial and error, as one needs to fine-tune hyperparameters and architectures to the task at hand. After training, hospitals need to ensure that the AI-based models are available for use by doctors, and so need to compress these models for smaller mobile devices, ensuring cross-compatibility with these devices.

What are the considerations involved in your choice of graduate school?

The graduate school I attend would ideally have the resources necessary to support both my interest in research and my desire to teach. For research, my graduate school would need professors who specialize in AI for social good, applying techniques from AI to practical fields such as healthcare and social networks. An example of a school that performs this research is Harvard, where the Teamcore lab with Prof. Milind Tambe utilizes machine learning (ML) algorithms for graphs to analyze social networks. For one project, the team investigated HIV intervention policies by determining central nodes in a graph using deep learning-based algorithms. This was then deployed by working with Safe Place for Youth, an organization that provides resources for homeless youth, resulting in improved HIV testing rates.

AI for social good is inherently interdisciplinary, requiring input from fields including medicine and economics. Because of this, I aim to go to a graduate school that prioritizes interdisciplinary collaboration. An example of such a school is Columbia, where interdisciplinary projects include one where a team of computer scientists and biologists work together to develop robots for people with disabilities.

I additionally aim to teach and mentor students in graduate school. An ideal graduate school would not only have graduate teaching assistant programs but also mentoring programs to work with local kids, especially in underserved neighborhoods. Examples of such programs include StreetCode at Stanford and Middle School Bridge at CMU. StreetCode works with communities of color to teach basic coding skills and Middle School Bridge assists high school freshmen with math. These programs allow me to practice teaching while also giving back to my local community. In graduate school, I aim to perform high-impact research and also teach, so the graduate school I go to must have opportunities for both.

Provide a concise summary, similar to a resume, in chronological order, with dates, recapitulating significant periods of technical and other creative activity since high school graduation. Omit activities only distantly related to your professional development. Include workshops, summer schools, a general description of all courses of study pursued (e.g. "3 quarters of Differential Equations") and degrees expected or awarded (dates, institutions, fields). Separate your undergraduate activities from your graduate activities (if/as applicable) with a single dashed line.

----2018----
----September---
Started school at UMD, ACES Honors College for cybersecurity
Worked in Aravind Srinivasan and Max Leiserson's lab to discover mutational processes through non-negative matrix factorization

---October---
Attended StringBio conference in Orlando on string algorithms for computational biology

----2019----
---January---
Started as TA for class on web programming, CMSC122

---May to August----
Worked at Carnegie Mellon University for the REUSE REU program with Bogdan Vasilesuc and Christian Kaestner on detecting online toxicity in software engineering communities.

----September---
Started working with John Dickerson on rideshare fairness
Started working with Jordan Boyd-Graber on improving entity linking through data collection
Started as TA for class on programming languages, CMSC330

---2020---
---January---
Published work at Carnegie Mellon, Raman on toxicity, Naveen, et al. "Stress and burnout in open source: Toward finding, understanding, and mitigating unhealthy interactions." Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: New Ideas and Emerging Results. 2020.

Started as TA for class on coding interviews, CMSC389O

----March---
Presented work on entity linking at MASC-SLL conference on speech and language

----May-August----
Worked at Facebook as a software engineer, developing web UIs for their feed ranking tools

---November----
Published work on entity linking with Jordan Boyd-Graber at HAMLETS workshop at NeurIPS
Published work on rideshare fairness with John Dickerson at Machine Learning for Econ Policy at NeurIPS
Published work on rideshare fairness with John Dickerson at AAAI Undergraduate Consortium

---2021---
---January---
Became head facilitator for student-taught class on coding interviews (CMSC389O)
Proposed new project on improving pricing systems for rideshare systems by including network externalities

---January-May---
Took a class on computational linguistics, working on a class project on bias in question answering, published in September at Machine Reading for Question Answering at EMNLP 2021

---April---
Published work on rideshare fairness with John Dickerson at IJCAI 2021

---May-August---
Worked at Lincoln Labs with Michael Yee on improving human-AI collaboration through fine-tuning and semi-supervised learning

---September---
Started working with Partha Lahiri on improving edit imputation algorithms

--October--

Published work done at Lincoln Labs with Michael Yee at Workshop on Human-Machine Decisions, NeurIPS 2021

---Course Highlights Major----
1 semester in following computer science graduate classes: Computational Linguistics, Deep Learning, Mechanism Design, Quantum Computing, Computational Geometry
1 semester in following science undergraduate classes: Computer Vision, Cryptography, Databases, Algorithms 2
1 semester in following math classes: partial differential equations, real analysis, probability theory, statistical theory, numerical methods, advanced linear algebra
1 semester in following lower-level classes: programming languages, algorithms, computer systems, calculus 3, tested out of linear algebra

---Course Highlights, Non-Major---
1 semester in agent-based modeling, natural resource economics, educational policy, physical chemistry, naval ethics, reverse engineering
3 semesters in cybersecurity

Use this page to submit a personal essay that reveals your personal creativity beyond your research. You may choose to focus on activities since your graduation from high school, but these do not have to be related to the classroom or lab, to STEM or your future career—they can also be from your more informal pursuits and passions, e.g., your extracurricular and leisure activities. This is an opportunity for you to share your narrative with experiences that demonstrate your creativity, perseverance, resiliency, curiosity and/or grit. You may use more than the allotted space for your response, and any additional text will be included in your application. (~2 pages?)

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 artificial intelligence (AI) for social good arose after reading stories on the power of AI and its racial bias issues, while my interest in research came from a desire to tell my own stories. While 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.

I engage in two types of teaching: community-based teaching through the Maryland Mentor Program, and university-based teaching through being a teaching assistant (TA). Through the Maryland Mentor Program, I teach reading and math to local elementary school students, allowing me to connect with my community. For the program last semester, I was paired with a mentee and worked with him on reading each week. Reading is an especially tough subject to teach; understanding that letters can combine to make words is a challenging concept, only complicated by the preponderance of arcane phonetic rules. To get around this, each week I would try different strategies to help my mentee understand sound blending: one week, I had him say letter sounds quickly, while in another I used rhyming words to show him similarities between "bat" and "cat." The latter technique worked particularly well, as my mentee was able to isolate the "b" and "c" sounds. I worked with him to build on this technique, using these sounds from rhymes to build words, which helped him incrementally progress in his reading abilities.

At my university, I teach two classes: coding interviews and programming languages. I am the head teaching assistant (TA) for the student-taught coding interview class, meaning I'm in charge of producing all content for the class and managing 5 sections of a 150 person class. This requires me to develop lecture videos, iterate through feedback from prior semesters to improve lessons, and develop homework. After talking with students, one thing I learned was that people tended to struggle with behavioral interviews due to their nervousness. To assist students with this, we introduced a new lecture on behavioral interviews, for which I developed slides, a lecture video, new homework, and practice problems for in-class participation. Similarly, because of the increase in take-home programming assignments, I developed a lesson on requests and APIs. Because our class focuses on

coding interviews, the midterm and final is a mock interview with a course facilitator. While running the mock interviews is difficult, as you need conceptual mastery to be able to help students, teaching is rewarding, as it's nice to see the "ah-ha" moment when a student understands a difficult topic.

Outside of teaching, I engage in activities, such as Quizbowl, puzzling, and botany, that explore my curiosities. These activities arise from an interest in discovering new stories and combining ideas from stories together. Participating in Quizbowl, a team-based academic competition similar to Jeopardy, allows me to read and discover new stories and topics. Quizbowl questions run the gamut from literature to science to pop culture, allowing me to understand topics outside of my STEM and computer science focus. For example, after hearing a question about *Go Tell It On the Mountain* by James Baldwin (a novel from the 1950s), I was inspired to read it, learning about life during the Harlem Renaissance. The story was especially interesting because of the empathy I felt for the main character (John Grimes) despite sharing little similarities, allowing me to explore a different viewpoint. Quizbowl encourages me to expand my reading palette, leading to an increased awareness of different perspectives.

Studying for Quizbowl does not occur overnight, and is rather the culmination of years of experience. Because of this, I aim to cultivate knowledge through taking diverse classes, especially outside of my major. A good example of this is the physical chemistry class I'm taking this semester. My lack of a formal chemistry background makes it difficult to understand certain concepts, such as the degrees of freedom for a molecule, and so to counteract this, I read through the supplemental material to gain a better understanding. The class is not only useful for Quizbowl, but is also interesting in its own right; we recently learned about the idea of negative temperature (where materials can decrease energy with increasing entropy), which is a paradoxical yet interesting concept that I would never have learned about had I stuck to my major classes.

Creating puzzles allows me to combine the stories and topics I've learned about from Quizbowl in an interesting, and eccentric, manner. As part of my puzzle club at school, I help design and run puzzle hunts, which consist of sets of puzzle challenges completed in teams against a ticking clock. These puzzles are open-ended, allowing me to explore various, seemingly unrelated topics. One of my favorite puzzles involved semaphores, which are flag-based signals, combined with country flags and NATO codes (which I looked into because I was fascinated with geography). My goal was to clue in on the various interesting flag patterns I had learned about through Quizbowl, and creating the puzzle allowed me to learn more about geography and semaphores in the process.

Designing puzzles is a matter of trial and error, as many times, the first puzzle created lacks polish or is too difficult. One of my most esoteric puzzles required participants to create a map involving a collage of streets from various cities. Designing the puzzle itself was hard, as I had to work backward from an intended answer to decide on streets and ways to clue at them. After testing the puzzle with others in the club, I found it was too difficult, so I spent hours adding in extra clues and

working with puzzle testers to ensure that it was solvable. During the puzzle hunt, I was nervous that the puzzle would be unsolved, but was excited when multiple teams managed to solve it.

My interest in botany arose from a fascination with plants; I was curious how and why small seeds could grow into large plants. Rather than finding the answers through books and stories, I decided to try hands-on experiments and grow plants. To start, I bought a set of radish plants, as they germinate within a few days, and grew them inside my apartment. I tried various containers for plants, including plastic, glass, and metal cans. I watered these plants daily, maintaining observations in a notebook. While some plants perished (the one in the plastic container died within the first week), others persisted for months. I keep re-running experiments with these plants, varying the amount of water given, amount of seeds used, and other quantities, to try and see what helps radishes grow the best. While growing plants is tough and requires patience, it's rewarding seeing them grow from little seedlings to full plants. I hope that my plants lead to edible radishes one day.

The skills I've learned through my activities are incredibly useful for research. Science takes patience and creativity; experiments rarely work on the first iteration and require creativity to iterate experiments based on prior failures. By engaging with activities such as teaching, Quizbowl, puzzling, and botany, I've learned these perseverance and creativity skills while exploring areas outside of my major.

Academic Honors

Honors/Fellowships

CS Honors College - Research-based honors college, allowing students to take classes on research, create a senior thesis
Global Fellows Program - Honors program that explores international and policy issues, culminating in a part-time internship
ACES Honors College - Cybersecurity-based honors college that teaches basics of cybersecurity, Linux, bash scripting. Developed cyber security-based research project
CMU REUSE REU Fellowship - Participated in a research program at CMU for 10 weeks, researching toxicity in online communities.
Math Modelling - Received outstanding award for modeling of refugee populations for SCUDEM math modeling competition
Bloomberg Codecon - Received top 3 awards at University of Maryland programming competition, and qualified for national competition at Bloomberg headquarters in New York City

List here, in chronological order, any fellowships, scholarships, teaching or other appointments held since entering college or university. Separate your undergraduate and graduate education intervals (if/as applicable) with a single dashed line. Put an asterisk (*) at the beginning of all lines indicating a national-level award (e.g., award of a National Merit Scholarship; NSF, NASA, or NDSEG Fellowship; election to Phi Beta Kappa, Tau Beta Pi, or Sigma Xi; etc.). Use one line per item whenever feasible (provide basic what, where, when data). You may use more than the allotted space for your response, and any additional text will be included in your application.

Scholarships

Goldwater Scholarship - Prestigious scholarship given to top undergraduate researchers nationally
Philip Merill Presidential Scholarship - Given to top 15 seniors for academic and service record
Iribe Scholarship - Computer Science scholarship worth 11K for academic and research record, given for 2 years (11K each year)
Presidential Scholarship - 4-year scholarship worth 20K given to rising freshmen
Capital One Scholarship - Computer Science scholarship worth 1K for academic and research record
Corporate Partners Scholarship - Computer Science scholarship worth 2K for academic and research record

Teaching

CMSC330 TA (Programming Languages) - Fall 2019-Present
CMSC389O TA (Coding Interviews) - Spring 2020 - Present, head TA Spring 2021-Present

CMSC122 TA - Spring 2019
Maryland Mentors Program, volunteer at local elementary schools - Fall 2020 - Present
College Park Academy Volunteer (volunteering at local charter school) - Fall 2018 - Spring 2021, assisted with cybersecurity club, math tutoring
College Mentors (helping middle schoolers with career pathways) - Spring 2021

First, please list the most significant research projects that you have pursued, in chronological order. (Include reference information for those that have been formally documented, presented at a conference, or submitted for publication.)

1) A Muffin-Theorem Generator at Fun with Algorithms (FUN) 2018, PI: Bill Gasarch, May 2016-August 2016
2) Discovering mutational signatures through non-negative matrix factorization, PI: Aravind Srinivasan and Max Leiserson, May 2018-August 2019
3) "Stress and Burnout in Open Source: Toward Finding, Understanding, and Mitigating Unhealthy Interactions" at ICSE 2020, PI: Bogdan Vasilescu and Christian Kaestner, May 2019-August 2019
4) "Investigating methods of balancing inequality and efficiency in Ride Pooling" at AAAI Undergraduate Consortium 2021, PI: John Dickerson, September 2019 - January 2021
5) "Data-Driven Methods for Balancing Fairness and Efficiency in Ride-Pooling" at ML For Economics Policy NeurIPS 2020 Workshop, PI: John Dickerson, September 2019 - January 2021
6) "What more can entity linking do for Question Answering at HAMLETS" NeurIPS 2020 Workshop, PI: Jordan-Boyd Graber, September 2019-Current
7) "Data-Driven Methods for Balancing Fairness and Efficiency in Ride-Pooling", presented/published at IJCAI 2021, PI: John Dickerson, September 2019 - January 2021
8) "Eliciting Bias in Question Answering Models through Ambiguity" at MRQA EMNLP 2021 Workshop, PI: Jordan-Boyd Graber, January 2021-August 2021
9) "Improving human-AI deference algorithms through fine-tuning", PI: Michael Yee, done while at Lincoln Labs, May 2021-August 2021, Workshop on Human-Machine Decisions at NeurIPS
10) "Improving fairness in rideshare dynamic pricing algorithms", PI: John Dickerson, January 2021-Present, currently under submission at AAAI Student Poster/Abstract session
11) Improving data imputation methods using machine learning, PI: Partha Lahiri, in progress

Finally, choose one or two projects that best exemplify your own creativity and discuss in more detail what you personally contributed to them.

Fairness in ride-pooling: I worked with Professor Dickerson to develop fairer ride-pooling matching algorithms, which match riders and drivers for companies like Uber, to ensure algorithms service neighborhoods equitably and minimize driver income inequality. Optimizing for fairness in ride-pooling is difficult due to the combinatorial complexity of rider-driver combinations, and so requires approximate solutions. To solve the problem, I first formalized different notions of fairness by developing fairness-based objective functions. To evaluate these objective functions, I extended a Markov decision process framework that matched riders and drivers by generalizing the reward and value function. By modifying the value function, I was able to tailor the value iteration process for the specific objective function, allowing for long-term maximization of fairness. After approximating the value function through deep learning, I incorporated the value function into the weights for a linear program. I found that these new objective functions increased pickup rates for riders in underserved neighborhoods while minimally impacting profit. While the new objective functions improved rider-side fairness, I realized that the objective functions did not manage to reduce driver-side inequality, and so I worked with a graduate student to develop an income redistribution method. This method uses the Shapley value to determine the expected earnings of a driver and redistributed depending on the Shapley value of drivers. I proved that the redistribution system guarantees a minimum wage, and found that the redistribution system reduced wage disparities while avoiding the free-rider problem. I worked with Professor Dickerson to write up our work, which I published and presented at the *ML For Economics Policy Workshop at NeurIPS 2020*, *Undergraduate Consortium at AAAI 2021*, and *IJCAI 2021*.

Human-AI Collaboration: This summer (2021), I worked with Dr. Yee at the MIT Lincoln Labs to improve human-AI collaboration. Learning-to-defer algorithms determine how to partition tasks between AI and humans, but are based on aggregate human performance rather than being fine-tuned for individuals. Improved learning-to-defer algorithms can improve system performance for autonomous vehicles, saving lives from car accidents. I developed the project from scratch by reading through prior literature to find weaknesses in AI, and I additionally reached out to other researchers at Lincoln Labs to discuss potential avenues for the project. To start the project, I developed a baseline fine-tuning algorithm that can be incorporated into learning-to-defer algorithms, then improved them using a self-training algorithm, which is necessary due to labeled data sparsity. The self-training algorithm adopts ideas from semi-supervised learning to annotate unlabeled data points and train using them. To evaluate these algorithms, I developed a synthetic dataset from scratch based on autonomous driving and showed that fine-tuning reduces driving time. I worked on deference in the CIFAR10 image classification task by analyzing human performance through dimensionality

reduction algorithms like PCA. I then used deep ResNet models to predict human performance and extended prior work by combining two fine-tuning algorithms with a modified cross-entropy loss function. Our work was published at the *Human-Machine Decisions Workshop at NeurIPS 2021*.