When I sit at the piano, my music flows naturally as my imagination leads my hands, carried by years of technical practice and the pursuit of a Bachelor's in Piano Performance. Music also inspired my approach to CS — starting from an intuitive picture, working through technical grinds, and arriving at a precisely formulated argument.

I have broad interests in theoretical computer science and have approached it through the **foundations of modern Machine Learning** (ML) and **Algorithmic Game Theory**. In both, I aim to understand the limits and capabilities of complex computational systems.

- (**Foundations of Modern ML.**) *For algorithmic tasks, when and why do Transformers prioritize learning statistical heuristics over generalizable algorithms, and how can we systematically induce robust reasoning?*
- (**Algorithmic Game Theory**.) *In collective decision-making, simple tournament-based voting rules are theoretically incapable of achieving optimal social welfare on their own. How can we bridge this optimality gap with minimally added overhead?*

Foundations of Modern ML. While Transformer models demonstrate remarkable capabilities, they often prioritize learning statistical shortcuts over generalizable algorithms. Advised by Professors Robin Jia and Vatsal Sharan, I investigated this failure on the graph connectivity problem and proved that the model's failure to generalize is not inherent to the architecture, but rather a deterministic consequence of the training distribution relative to the model's capacity. Crucially, this implies that by **aligning the training distribution with the model's theoretical capacity**, we can steer models toward more **robust out-of-distribution generalization**.

I developed the theoretical foundations of this project. I first proved a tight, non-asymptotic capacity bound showing that an L-layer model has the exact circuit complexity to solve connectivity for graphs with diameter up to 3^L. This theoretical threshold allowed us to model the learning process as a competition between two latent channels: a robust "algorithmic" channel that implements matrix powering and a shortcut "heuristic" channel that relies on a degree-counting shortcut. I then identified a sharp phase transition that depends solely on the data composition: "within-capacity" graphs drive the gradient descent trajectory toward the algorithmic solution, while "beyond-capacity" graphs promote the heuristic. This finding offers a prescriptive strategy for out-of-distribution generalization: restricting training data to graphs within the model's capacity paradoxically encourages the model to learn the generalizable algorithm by suppressing the heuristic channel, a prediction my collaborators empirically validated.

Algorithmic Game Theory. Under the metric distortion framework, the design of voting systems involves a fundamental tension between simplicity and efficiency. Tournament-based voting rules are practically appealing because they minimize cognitive load, requiring voters to compare only two options at a time. However, this simplicity comes at a theoretical cost: standard tournament rules are provably suboptimal compared to the general deterministic optimum.

Advised by Professor Kamesh Munagala, I sought to bridge this gap while preserving the simplicity of tournament rules. We proposed "Deliberation via Matching," a protocol where voters with opposing views may engage in pairwise discussions. We proved that by augmenting tournament rules with minimal pairwise deliberations, we can break the aforementioned barrier. Conceptually, this establishes that the cognitive simplicity of **tournament rules** does not require sacrificing social welfare: with slight overhead, they are just **as powerful as general deterministic social choice rules**.

In more technical detail: I led the theoretical analysis to resolve the analytical intractability that had limited prior work. The distortion objective for deliberative processes is inherently non-linear and non-convex, forcing previous studies to essentially rely on black-box numerical optimizers. I identified that the objective is *supermodular* and *convex* with respect to the metric variables. These were the crucial keys to tractability. Supermodularity implied that worst-case instances must follow a specific monotonic "structure," while convexity allowed me to apply Jensen's inequality locally to collapse a continuum of voters into a discrete set. This reformulation converted the intractable program into a bilinear relaxation, from which the optimal solutions easily followed via vertex enumerations and linear programming. While the proof is terse in its algebra, each step closely follows an intuitive story, and this precisely echoes my belief that exciting research can start from and be driven by clean intuitions.

At Harvard, I hope to contribute to the **EconCS group** and work with Professors **Ariel Procaccia** and **Yiling Chen**. **I am eager to explore the vast landscape of Algorithmic Game Theory**. In reading recent work at Harvard EconCS, I frequently return to a common conceptual theme: *mechanisms not only aggregate preferences and information, they also shape what gets expressed and who participates*. I want to study how to design and analyze mechanisms that remain principled under that feedback, and I see two specific, complementary perspectives to pursue this goal.

First, I am drawn to Professor Procaccia's recent research that treats deliberation and representation *as algorithmic objects*, from public-spirited voting (EC'23), where he showed that for some rules, even a small degree of public-spiritedness can turn unbounded (utilitarian) distortion into constant, to citizens' assemblies under attrition (EC'25), and to (the preprint on) auditing justified representation. This line of work interests me because it puts institutional bottlenecks (e.g. behavioral assumptions) directly into the model, making the research question highly practical, all while doing so via theoretical approaches.

Professor Chen's work supplies a complementary view, where *information constraints* become part of the mechanism. I am drawn to this perspective as it makes a recurring implicit assumption explicit: the procedure only gets whatever evidence the setting can stably generate, and that evidence is shaped by incentives and strategic adaptation. Her recent works on reference-agent scoring (EC'25), where agents can be carefully scored even when no ground-truth outcome avail, as well as on bonus-penalty elicitation for comparisons (NeurIPS'24) and equilibrium-aware strategic classification (ICLR'25) all give clean instantiations of this agenda.