When Can (and Should) AI Represent Student Preferences?

By Erica Chiang (Cornell Tech)

When Can (and Should) AI Represent Student Preferences?

By Erica Chiang (Cornell Tech)

A theme that has emerged in our DLI working group so far is the idea of how AI can be used to represent complex aspects of human values, identities, and preferences. It quickly became clear that depending on our individual experiences, we were either using the word "can" to reason about what AI can *technically and practically* achieve, or to reason about what AI *should* be used for from an *ethical perspective*. Our group has worked to develop a framework that places AI applications along these two axes (technical feasibility and moral implications) to explore how these dimensions can diverge.

In this critical reflection, I use this framework to examine a project I am working on with New York City Public Schools (NYCPS) to nudge certain students toward high-performing high schools. While the project is motivated by a clear normative goal—reducing information gaps in the high school application process—it surfaces deeper questions about when models should attempt to represent student preferences, and what responsibilities arise when it does.

Nudges to high-performing high schools to address disparities in high school application behavior

In New York City, eighth-grade students apply to high school by submitting a ranked list of programs they are interested in attending. A centralized matching process then assigns them to a single program where they receive an offer. This system was designed to expand access to schools by allowing students to apply to any high school in the city and to rank programs based on their true preferences (without needing to strategize based on their chances of admission).

But in practice, the scale and complexity of the system makes this difficult. NYC has over 900 high school programs, with a wide range of academic focuses, extracurricular activities, and performance metrics, meaning that forming informed preferences over schools requires significant time, information, and support. This complexity may be a driving force behind disparities in high school application behavior—for instance, we observed that students from some middle schools consistently apply to high-performing high schools at much lower rates, even to schools where they are likely to be admitted. If preferences are shaped by informational barriers, then using AI to model student preferences could help students with preference discovery.

With NYCPS, I designed an informational intervention to target students from middle schools that have not historically sent many students to high performing high schools, helping them discover **high-performing, nearby high schools where they have a strong chance of admission**. Our goals were to (1) expand opportunity and increase awareness of high-quality school options for students who might not otherwise have known of them, and (2) understand the extent to which current behavior reflects true preferences versus lack of information. Designing the intervention required close collaboration across multiple NYCPS offices, as well as focus groups with families and counselors. These conversations shaped practical design choices such as how far a recommended school could be from a student's home, how many schools to recommend, and how to communicate recommendations in a way that was informative without being prescriptive. These conversations engaged critically with both the technical feasibility and the normative goals and ethical implications of the project, and through the project's evolution we tried to maintain clarity on where our project sat along both axes.

System-level considerations

A critical aspect of the intervention was **anticipating system-level effects**. Because the match is centralized and capacitated, nudges could inadvertently create congestion and

change program acceptance rates, with risks of exacerbating disparities. This revealed an important tension in our framework. Despite the intention of the project to assist students and reduce information gaps, overlooking downstream effects of technical success in one stage (identifying students to nudge) could ultimately end up negatively impacting students.

Based on this insight, I used historical data to simulate match outcomes under different adoption scenarios to determine the maximum safe number of nudges per program. This allowed us to expand access without significantly affecting admissions cutoffs, ensuring that the recommendations respected existing operational and ethical constraints.

Further, even upon realizing that addressing congestion and system-level impacts were key design goals of our intervention, it became clear that **technical failure** itself could have ethical consequences. If the system failed to handle congestion properly, the intervention could inadvertently hurt students, including those it aimed to help. This highlighted an interesting relationship between the two axes in our framework: while researchers should look beyond a project's technical scope and consider its broader social impact, the technical ambition and feasibility of a project are still crucial because technical failure can directly shape the consequences of deploying a system.

User agency

Finally, when deploying the system to about 500 eighth-grade students currently applying to high school (a number chosen to be small enough that we could develop confidence in the system), we worked extensively to design language that emphasized to students that they have full control over their decisions and that did not imply any schools were better than others.

This reflected an important recognition: while a technical system can help students navigate a volume of programs that they would not be able to otherwise and mitigate

disparities in information access, it cannot—and should not—claim to perfectly represent any student's true preferences. Preferences are shaped by values, identities, family considerations, and lived experiences that extend far beyond what our technical capabilities allowed us to capture. Acknowledging this limitation and explicitly designing around it was crucial for preserving student agency and trust.

Conclusion

This project reinforces the importance of situating technical interventions within both technical and moral frameworks. While technical systems can help surface options and approximate aspects of human preference, their use in complex social systems demands sustained attention to equity, system-level dynamics, and user agency. Within one technical project, the ethical impact is shaped not only by the high-level goal, but by the details of how it is implemented, constrained, and communicated.

Erica Chiang
DLI Doctoral Fellow
esc99@cornell.edu
More Info >