Ruru Hoong
ruruhoong@g.harvard.edu
Course of graduate study: Business Economics at Harvard Business School / Department of Economics (enrolled for 3 years).

My interests in digital economics and AI informs most of my current work and are three-fold: (i) assessing the heterogeneous impacts of AI-driven technologies, particularly on educational outcomes; (ii) understanding the technological shocks to, and labour market structure of, the AI data annotation industry, and (iii), understanding how to utilise ML methods to estimate causal impacts of interventions in my own applied work.

My main current research project concerns the impact of AI technologies on education (AIEd). Developments in ML models in personalised-instruction hold promise in increasing access to learning, equalising the learning process, and reducing the costs of human capital acquisition. Does the educational experience generated by ML-personalised predictive technologies help close or exacerbate the achievement gap? Through what channels do they help? Given that new data from any additional or existing student updates the prediction algorithm, additional information helps improve predictions for the students that come after. Since ML models depend on training data, the order in which students and schools are recruited to participate in such technologies matters towards the efficacy and equitability of the product – for example, rolling out to highly-motivated students and teachers first as opposed to going in a random order might make a difference. When forming predictions for new students, it is possible that the model may be more well-tailored to students who are more similar to existing students in the platform. By collaborating with a large AIEd firm, I aim to (i) evaluate the heterogeneous impact of AIEd on student achievement, (ii) understand the channels through which adoption of such a technology affects outcomes, and (iii) measure the spillovers of early adoption of the technology on those who adopt it after.

To study some of the questions posed above, I'm currently analysing the historical data of a Seoul-based AIEd platform that administers personalized prep questions to 4 million students taking the TOEIC professional English-language proficiency exam. The first question I am studying is how the order of platform rollout across localities affects its impact. This is a unique context to study the way in which education innovation – which often starts in rich countries and is then implemented in less well-off localities – impacts regional disparities in outcomes. Moreover, even within each of the countries, the adaptive nature of the technology differentially affects who benefits from it. Students that were previously the "target instruction-level" of courses in each locality might benefit less as personalization has less additive benefit for them; on the other hand, students who are "rare" in the data would not be well calibrated in the recommendation model. Who benefits and why would matter a lot depending on the incentive structure in place in each country, as well as differential rollout over time.

Second, I am studying how the acquisition of more labels – the outcome that the model is trained to predict – can improve educational efficacy, and whether or not it improves outcomes for some demographics versus others. The company has 10k labels attained through offering incentives to students to report their actual standardised score outcomes. How representative are the labelled users (e.g., the most well-performing students might be more willing to report their scores), and what does that mean for the performance of the model for different types of students? I am developing observational and experimental methods to measure how additional labels might affect efficacy, giving a quantifiable measure of "unbalanced dataset" bias.

Lastly, it could be fruitful to run a randomised control trial in tandem with the companies' expansion efforts to the SAT/ACT in the US to measure the causal effect of their products. A RCT would be a gold-standard evaluation; moreover, randomising over the schools/states in which rollout happens first can help answer some of the questions posed previously. This project could also speak to a broader question of "crossing the chasm" – how firms should take a product that is successful with early adopters to the wider mainstream market. On the one hand, for the reasons mentioned above, products driven by ML models might be biased towards early adopters; on the other hand, readaptation may come at a lower cost than to traditional firms.

A related question I have been exploring lies further upstream: in the production of data. With the fanfare over GPT4 and other foundation models, recent public discourse about AI models has proliferated. But little attention – nor academic work – has focused on what is arguably the most important input into these models: data. I have recently started working with an AI data annotation company to see how technological shocks may affect impact (i) the marginal product of GPU vs. human labor, (ii) wage structures, and (iii) how best to hire and structure work in the industry. Both labour as well as computing input matter towards the production of labelled data, but the composition and interaction of these inputs are constantly changing in a rapidly evolving industry. Technological changes have the potential to affect the nature of work and QA process. For example, some technological advances – like Meta's Segment Anything Model – may affect only one type of labelling (e.g., computer vision), affecting the nature and composition of jobs and payment structure in that domain but not others. This can also affect the optimal wage-structure: pay-by-piece/hour, or some other model – by working with a large data annotation firm's historical data across contracts and workers, I hope to shed light on some of these questions.

I am also furthering my work on digital data and AI by exploring the main policy instrument over the market for digital data: privacy regulation. In joint work with Andrew Kao and Professor David Yang, I've been working on a large-scale survey and experimental elicitation of privacy preferences and valuations of data privacy across the world. Through a series of surveys and experimental methods, we aim to empirically measure, describe, and understand citizens' preferences for privacy around the world, which could shape data-related regulations and the supply of data, a key input in the data-intensive economy. Preferences for privacy can differ depending on the structure of the political institutions that operate in different countries, as well as the perception people have of private firms as opposed to the government. We aim to shed light on within and cross-country heterogeneity in preferences. In parallel, we are also in the process of coding up historical privacy regulations across the world and looking at the impact of their adoption on AI output. Some preliminary results show that some aspects of privacy regulations (e.g., extraterritoriality) do have a potentially large impact on AI innovation.

Lastly, I have been interested utilising ML methods to estimate causal impacts of interventions in my own applied work since my undergraduate years. Based on Athey and Wager's (2019) application of causal forests to the National Study of Learning Mindsets dataset, I applied causal forests and a X-learner strategy to estimate treatment heterogeneity in my undergraduate thesis, which sought to understand digital addiction in the social media space. Under the guidance of Professor Matthew Gentzkow, I'd designed and implemented a randomised intervention, encouraging 600+ participants recruited over Facebook ads to adopt a voluntary soft commitment device (iOS "Screen Time" app limits) to limit their phone and

social media usage. My paper won the Firestone Medal (awarded to top 10% of theses university-wide) and was published in the European Economic Review. Given that I ran a randomised control trial, both strategies yielded unbiased estimates of the average and group-specific treatment effects. However, much of my future work might not be experimental. Through the summer school, I hope to further my understanding of the new methods used in ML and issues surrounding how they can be used for causal inference in settings with quasi-random variation (e.g, IV, RD designs), panel data settings, or structural models of consumer or firm choice.

I've continued this work on digital addiction in grad school: arising from my undergrad thesis is the following puzzle – if commitment is so widely available and social media users are so willing to take-up commitment when it is offered to them in experiments, then why is the natural use of commitment in the general marketplace so low? To answer this question, I sought a collaboration with a large social media platform in joint work with Peter Hickman and Yulu Tang. We document the demand for, and effectiveness of, soft commitment devices through a naturalistic experiment on 13 million active social media users – orders of magnitude larger than any previous study on commitment. We also aim to study the trade-off between short-term engagement and long-term satisfaction: in order to do so, we are applying ML methods to predict (at baseline) how likely our 13 million participants are to uninstall the app in the next six months, as well as their daily usage over future periods. Studying heterogeneous treatment effects of soft commitment devices using these as moderators could then shed light on this trade-off: if platform uninstallment is truly a substitute for soft limits, then those ex-ante predicted most likely to uninstall the platform should see the highest treatment effects.

I would love to continue exploring these topics, and more. It would therefore be a joy and an honour to participate in the Digital Economics and AI tutorial, not only to learn directly from the leading researchers in this space, but also to meet and develop relationships with other students and faculty interested in questions such as these.