

# Discussion of “Addressing the Challenges of AI-Generated Assignment Submissions in Education: Insights and Strategies”<sup>☆</sup>

STEPHANIE L. CHAN<sup>1</sup> AND ANDREW ADRIAN Y. PUA<sup>1,\*</sup>

<sup>1</sup>*Carlos L. Tiu School of Economics, De La Salle University–Manila, The Philippines*

**Keywords** *academic integrity; assessment; ChatGPT; generative AI in education; higher education; learning*

We thank Wang et al. (2025) for organizing and providing 6 assessment strategies meant to ensure academic integrity in the deluge of AI-generated submissions. We have to be extremely adaptive and tactical with respect to our teaching and assessment strategies, but the essential minimum to contemplate is making it worth faculty time to check and assess student submissions, while making sure that students actually learn what we intend them to learn. Therefore, we think that the 3rd and 6th strategies proposed by the authors – designing open-ended, creative questions and encouraging human-AI collaboration – are the most feasible and practical, but we share experiences which show the potential limits of these strategies.

**Why Maintaining Academic Integrity in the Age of AI Is Tricky** Using AI detection software will inevitably be circumvented by artful individuals. On the other hand, an approach that promotes ethical use, like the 5th strategy proposed by the authors, relies on the state of the student before they get to use AI. If the focus of the program or the culture is “getting ahead”, then ethical training will likely have little effect. What might be more useful is to inculcate a strong sense of ownership of ideas into the students. How can students be expected to take responsibility for AI-generated output if they are unable to determine where AI ends and where their own thoughts begin?

The first 2 proposed strategies of incorporating oral examinations and presentations and requiring process documentation can be adaptive and tactical, but they do not scale well in terms of faculty time and effort. For a class size of 40, this strategy would take at least 30 minutes per student which amounts to at least 20 hours of assessment. Requiring process documentation by assessing logs of AI usage is time-consuming, just like checking supplementary appendices of research papers. Students can go through multiple iterations of AI prompting and only show the final set of prompts and responses, akin to presenting only statistically significant results.

The 4th strategy of leveraging AI for assessment support will only be useful for certain types of courses. How well a student understands the material would be harder to assess automatically with AI tools. It is too early to tell whether AI will eventually be able to replace assessments traditionally done by faculty.

**Experiences of the First Author** She recently taught intermediate microeconomics to undergraduate students. She conducted a survey eliciting preferences of students over different bundles of pencils and tissue packets, with and without budget constraints. Groups of at most

---

<sup>☆</sup>Main article: <https://doi.org/10.6339/25-JDS1208>.

\*Corresponding author. Email: [andrew.pua@dlsu.edu.ph](mailto:andrew.pua@dlsu.edu.ph).

4 students were then asked to draw indifference curves using actual preferences, to construct the implied demand curves from how the students used up the hypothetical budget, and to determine the extent to which actual preferences adhered to utility theory.

The project required manually plotting indifference curves, by assessing each new bundle against a reference bundle. This task was relatively AI-proof, as it would require that students be able to give clear instructions and to describe the data collected to AI. Many groups found graphical representation difficult based on the number of questions she received and the number of incompletely labeled graphs submitted. The groups found it easier to plot demand curves.

When asked about the consistency of their resulting demand curve with utility theory, many groups stated that diminishing marginal utility can be inferred from their graphical representations. Unfortunately, they did not provide any support for this statement. There was no explicit utility function provided by the first author, and there was no attempt from the groups to recover the utility function from the indifference curve. Upon seeing these results, the first author asked ChatGPT the same question and found the statement regarding diminishing marginal utility. This points to the danger of students taking the answers given by AI engines at face value, without looking for further proof.

**Experiences of the Second Author** He gave the following examination question to economics graduate students taking a probability and statistics course. What is the distribution of  $Y = -\log X$  if the random variable  $X$  has a pdf  $f_X(x) = 3x^2 I_{[0,1]}(x)$ ? In the closed-notes version of the exam which was taken first, the students only had access to a rolodex of distributions (which includes the beta distribution). In the AI-assisted version of the question which was taken *a week later*, he asked students to simulate 10000 random draws directly from  $f_X$  and to provide graphical evidence that the resulting distribution of  $Y$  is one of the parametric distributions in the textbook. What did he find?

He found that most students in the closed-notes version were able to show that  $Y \sim \text{Exp}(1/3)$  as the rolodex uses a parametrization in terms of the scale rather than the rate. In the AI-assisted version, all students *except one* provided AI-generated R code which uses inverse transform sampling via  $U(0, 1)$  rather than using the beta pdf. They provided graphical evidence that  $Y \sim \text{Exp}(3)$ , which did not follow the parametrization in the book.

Traditional questions with open-ended elements can be useful for assessing students, especially where the gaps are. Students did not add value highlighting their discerning use. As a result, they did not present their solutions in a concise fashion, did not efficiently exploit the resources available to them, and did not do a sanity check of what was really requested of them.

**Concluding Remarks** Farrell et al. (2025) have emphasized that “large language models should not be viewed primarily as intelligent agents but as a new kind of cultural and social technology, allowing humans to take advantage of information other humans have accumulated.” The instant answers make the payoff to discovery small, therefore, we need to convince students that they have a stake in contributing to this technology.

It is easy to be swept up in the AI trend, but that in itself changes the way *everyone* engages with the material. We also have to contend with the decreasing attention spans. Perhaps we should look at AI as an invitation to change how we evaluate, how we deliver the material in the first place, and how we address our own gaps in knowledge, skills, and attitudes.

## References

- Farrell H, Gopnik A, Shalizi C, Evans J (2025). Large AI models are cultural and social technologies. *Science*, 387(6739): 1153–1156. <https://doi.org/10.1126/science.adt9819>
- Wang S, Xu L, Liu J, Zhai Y (2025). Addressing the challenges of AI-generated assignment submissions in education: Insights and strategies. *Journal of Data Science*, 24(1): 1–7. <https://doi.org/10.6339/25-JDS1208>