Assessing Students Amidst the Challenge of ChatGPT

Michael H. Wang[1,*] and Meihua Lee[2]

[1] Faculty of Engineering, University of Windsor, Windsor, ON, Canada
[2] Department of Communication & Technology, National Yang Ming Chiao Tung University, Hsin Chu City, Taiwan, China
Email: wang5@uwindsor.ca (M.H.W.); jenniferlee@nycu.edu.tw (M.L.)
*Corresponding author

Abstract—Recently, the rapid advancement of AI hardware and tools has led to the widespread adoption of natural language transformers like OpenAI's ChatGPT, Google Bard, Bing AI, and others in various business sectors. Nevertheless, for the academic community, these AI tools present both opportunities and threats. Like their counterparts in the business and industrial sectors, academics can leverage these AI tools for coding, idea/concept generation, planning, and other applications, benefiting from their global usage. However, the academic community also harbors concerns regarding the potential impact on academic integrity, as students may be tempted to rely on these tools to complete their essays, assignments, and exams without putting in their own efforts. In this article, we will present the authors' approach and findings in dealing with these AI tools while evaluating students' performance with two university student groups: engineering (Canadian) and communication technology (Taiwanese). We have identified key guidelines to deter students from directly copying answers provided by AI tools like ChatGPT. However, it is important to recognize that this approach will be an ongoing process, as AI tools continuously learn and adapt to new cases.

Keywords—Artificial Intelligence (AI), student assessment, ChatGPT

I. INTRODUCTION

ChatGPT, an AI-based natural language transformer aimed at developing human language responses to natural language queries, has emerged as a prominent subject of discussion in recent times. Its applications in the business community span through various fields such as coding, planning, advertising copy writing, and others. In the academic fields, researchers, professors, and teachers, as well as students all showed strong interests in adopting this new tool to facilitate their work. However, the growing interest in those AI tools has also sparked a range of ethical issues such as authorship, information reliability, genuine contributions of research endeavors, and specifically academic integrity for student performance assessment. This article endeavors to explore the ways that university professors can adopt when developing examination and assignment questions that can hinder the students from taking advantage of the AI tools, i.e., input the questions to the AI tool and generate answers within seconds.

Specifically, we are focusing on the ChatGPT (Ver 3.5 free version) language transformer. Other similar AI tools such as Google Bard, Bing AI are not demonstrated in this article as their developments are lagged behind ChatGPT, as of June 2023. Another point to ponder is that we are focusing on the free version of ChatGPT (Ver 3.5) in this study rather than the paid version (Ver 4.0) as we believe that most students are taking advantage of the free version instead of paying monthly fees for the advanced version. We also outline the main difference between these two versions. First, for the free version, users can upload up to 3000 words of content for specific training and learning purposes before queries; while for the paid version users can upload up to 25000 words of content. Second, the free version will not be able to provide description or interpretation of figures and graphs, while the paid version can provide minimal description of figures and graphs, even though the quality of those descriptions may not be at satisfactory level. As of July 2023, Google Bard also announced its ability to interpret pictures and figures [1, 2–9].

II. SAMPLE QUESTIONS AND ANSWERS GENERATED

We are presenting several examination questions from two student groups. The first group comprises engineering students from a Canadian university, and the second group is a group of students in the field of social science (communication technology) from a Taiwanese university. The questions cited are actual questions we gave to students in Engineering Economics, Statistical Quality Control, and Design of Experiments (Canadian group), and New Technology & Organizational Communication (Taiwanese Group). The ChatGPT answers toward those questions are not fully given, except for the social science subject (#6), as the length of the answers is quite lengthy.

Example Question 1 (Engineering Economics)

Q. Three credit card companies offer the following interest rates. Bank-A offers a 12% annual rate and compounded quarterly. Bank-B offers a 12% annual rate

but compounded monthly. Bank-C offers 12% annual rate but compounded weekly. Which credit card has the highest effective annual interest rate?

ChatGPT Answer: Correct

Discussion: This is a fairly straightforward question as the formula for compounded interest calculation is easy to comprehend. This type of question poses no challenge for ChatGPT or other AI tools.

Example Question 2 (Engineering Economics)

Q. Windsor Tool Inc. has a $500,000 loan for a new EDM machine to be used in tool/die production. The interest rate for this loan is 5% compounded annually. The finance manager decides that the company will make a $50,000 payment each year, starting the end of the first year. By calculation, it will take the company N years to pay back the loan. Notice that the payments for Year-1 to Year-(N-1) will be $50,000 as planned. The last payment at the end of Year-N will be smaller than $50,000.

First, calculate the value for N=? (years)

Second, calculate the last payment for Year-N=?

ChatGPT Answer: Incorrect

Discussion: The free version ChatGPT 3.5 completely missed the answers to this question. The content knowledge for this question is not considered as advanced in the engineering economics subject. However, the AI tool failed to calculate the proper interest payments and number of payments. When we provided ChatGPT with the specific lecture contents about this question and then queried about the response, the first question was answered correctly but the second part was still incorrect. If we specifically input the equations for both parts of the question, correct answers are generated.

Example Question 3 (Statistical Quality Control)

Q. A manufacturer of LED displays is interested in the effect of two different thin-film coating methods (Method-A and Method-B) on the thermal conductivity of the LED displays. The target LED thermal conductivity level is set as 1.0 (W/mK) and the specifications are USL = 1.50 (W/mK) and LSL = 0.70 (W/mK). An LED display with thermal conductivity outside the specification limits will be scrapped and will cost the manufacturer $100. Ten (10) LED displays are tested using each of the two methods and the results are as follows:

| A-met: | 1.20 | 1.10 | 1.00 | 0.90 | 1.20 |
| | 1.30 | 0.80 | 0.90 | 1.30 | 1.20 |

| B-met: | 1.40 | 1.10 | 0.80 | 0.90 | 1.40 |
| | 1.30 | 1.10 | 1.20 | 0.80 | 1.40 |

Using Taguchi's quality loss function concept, determine the quality levels for these two methods. Which method is better in terms of quality?

ChatGPT Answer: Incorrect

Discussion: This question poses some challenges for the ChatGPT as there are at least two levels of decision logic behind the question. First, the correct quality loss function should be selected based on different quality characteristics (N-Type symmetrical; N-Type Asymmetrical; Lager-the-better Type and Smaller-the-better Type). Second, after selecting the suitable quality loss function formula, the comparison between the two

methods (A and B) is to be calculated. In addition, when we present the question with tables or figures (for the two methods' data), this poses new obstacles for the ChatGPT as the AI tool needs to accurately process/interpret the figure/table containing the data. With the free version ChatGPT 3.5, this cannot be processed. Lessons learned here are to create a multi-level logic framework for your questions and provide some information and data in tables or best in pictures/graphs.

Example Question 4 (Design of Experiments)

Q. An industrial engineer is investigating the effect of four different component assembly procedures (A, B, C, D) on the quality characteristic of the printed wired boards. Four operators are selected for the study. Furthermore, the engineer has identified room temperature as another blocking factor which is set at four levels. The engineer uses the Latin Square design shown below (Table I) with the experiment results. Analyze the data from this experiment (using alpha = 0.05) and draw appropriate conclusions.

TABLE I. LATIN SQUARE DESIGN PARAMETERS

Room Temp. Level	Operator			
	1	2	3	4
1	C = 6	A = 7	B = 8	D = 11
2	A = 7	B = 7	D = 6	C = 8
3	D = 11	C = 8	A = 7	B = 10
4	B = 12	D = 10	C = 7	A = 7

ChatGPT Answer: Incorrect

Discussion: This question poses some challenges for ChatGPT. When the matrix table is presented as a table, the ANOVA table prepared by the ChatGPT is inaccurate, which is supervising. We found out that even though ChatGPT accurately calculates the means for all the factor levels, ChatGPT did not find the accurate Sum of Squares within the ANOVA table. We could not figure out the reason why ChatGPT cannot perform this calculation accurately, but we believe the main objective for those natural language transformer AI tools is not for statistical calculations. Similar to Example Question 2, when we present the data matrix as a figure/graph, ChatGPT cannot interpret those data and fail to answer the question completely.

Example Question 5 (New Technology & Organization Communication) (English translation is presented)

Q. (English) The scholars of 'The Montreal School' explore the Communicative Constitution of Organization (CCO) through linguistic, interpretive, and critical theories.

Please read [Report 2: Delivery Workers Face Wage Cuts Again/Revealing Various Forms of Exploitation, Lai Xiangling Criticizes Sweatshop Labor and Calls for Legislative Protection]. Using the conceptual perspectives of Text and Conversation, Co-Orientation, Distinction, and Scaling Up, discuss how the delivery workers' union organization can employ the six steps of scaling up: 1) The intent of the speaker is embedded in conversation; 2) The conversation is given a narrative representation; 3) The text is transcribed into a more permanent form; 4) A specialized language is developed that is used in

subsequent texts and conversations; 5) The texts and conversations are transformed into material and physical frames; 6) The standardized form is disseminated to a broader public. Apply these perspectives to the dialogue and text surrounding the 'delivery workers facing wage cuts,' engaging in co-orientation, distinction, and scaling up.

Answer provided by ChatGPT is considered moderate but lacks originality and context.

Full ChatGPT Answer: (Translation in English)

From the perspective of the Montreal School, we can apply the concept of organizational communication to analyze the dialogue and text surrounding the "reduction in wages for delivery workers" and expand and upgrade it through collective, extended, and amplified processes. Through the following six steps of expansion and upgrading, we can explore how the delivery workers' union applies these concepts to advance its goals.

1. Integrating the speaker's intention into the dialogue: Members of the delivery workers' union can begin expressing their concerns and dissatisfaction with the wage reduction issue in the dialogue. These conversations can include exchanges among workers, negotiations with management, and public statements.

2. Dialogues take on narrative forms: By transforming the content of the dialogue into narrative form, the delivery workers' union can convey the importance and impact of the wage reduction issue to a broader audience. This can be done through media reports, storytelling on social media, public speeches, and other forms.

3. Transcribing text into more durable forms: Transcribing the dialogue and narratives into written forms, such as reports, public statements, or declarations, makes them easier to preserve, disseminate, and review. These textual materials can serve as the basis for the union's subsequent actions.

4. Developing specialized language: At this stage, the delivery workers' union may start developing specialized terms, vocabulary, or symbols for use in future texts and dialogues. The development of this language can help members of the organization communicate their messages and demands more effectively.

5. Transforming text and dialogue into tangible and physical frameworks: The delivery workers' union can transform previous texts, dialogues, and relevant information into physical forms, such as posters, banners, and signs. These materials can be used for demonstrations, rallies, or other public events, further highlighting the importance of the wage reduction issue.

6. Standardizing and disseminating forms to a wider public: The delivery workers' union can standardize their demands and actions to more effectively communicate them to a broader public. This can be done through media coverage, social media promotion, website development, or other forms of public relations activities.

Through these six steps of expansion and upgrading, the delivery workers' union can gradually expand and upgrade the issue of "reduction in wages for delivery workers", gaining more attention and support from its members, the general public, and government authorities. This process

not only helps the organization foster collective alignment internally but also extends its influence to a wider social sphere, thus generating greater impact.

Discussion: We specifically presented the full answer provided by the ChatGPT about this question. The answer demonstrated the main difference between the type of questions involving logic and calculation (e.g., STEM subjects) and questions involving social science subjects where clear-cut answers are rare. The professor's assessment about the answer provided by the ChatGPT is moderate/acceptable but lacks originality and context. The answer looks more like a technical manual which explains and elaborates the process and stages given but fails to provide personal observations.

III. Lessons Learned from Sample Questions

Based on the demonstrated example questions and other tests we conducted, we have compiled the following guidelines for university instructors who aim to uphold academic integrity standards while assessing students through examinations or assignments, especially in situations where online tools cannot be restricted. Specifically, these assessment mechanisms, such as online course delivery and in-person assignment/tutorial sessions, are typical environments where these guidelines should be considered in the preparation of questions/problems, as online access cannot be disabled.

Guideline #1: Design questions with a mixture of texts, figures, graphs, and tables. When the question content involves more than just text, it complicates the task of natural language processing. Tools like ChatGPT may not completely and accurately capture the question content. In theory, all current AI tools like ChatGPT are trained with billions and trillions of data points, and their responses to specific queries are based on the most suitable and ranked answers, which may not always be entirely accurate. By creating obstacles like mixing different content types, it further prevents ChatGPT from accurately understanding the question statements, resulting in a lower quality of response. When combined with other guidelines discussed later, this approach can be highly effective in reducing the accuracy of capturing the question contents.

Guideline #2: Design questions with multiple logic levels (including multi-level 'what-if' scenarios or selecting the proper method/formula before proceeding to next step). As demonstrated in Example Question 3, when ChatGPT wasn't provided with the complete list of available models and formulas, it tended to use its learning mechanism to pick the best model, which may not always be accurate. This can be rectified by inputting all the formulas/models right before our query. However, students may not have the time and convenient access to the necessary material to feed the AI tool for generating the correct answer. Furthermore, students might need to provide more detailed information about the prerequisite conditions for applying a specific formula/model. When students can accurately identify the correct formula/model to arrive at a solution, they demonstrate mastery of some of the assessment objectives.

Guideline #3: Design questions with implicit information, such as natural and physical limits. While we did not demonstrate this guideline with the aforementioned examples, here's an exercise question we tested with ChatGPT: 'If it took Ben 2 minutes to drink a 1L bottle of Coke, how long will it take Ben to finish 500 1L bottles?' ChatGPT, not surprisingly, provided us with the answer of 1000 minutes. However, this answer is clearly inaccurate in real life, as human limits prevent a person from continuously consuming 500L of liquid. This guideline may be particularly relevant for Non-STEM subjects.

Guideline #4: When designing STEM-type questions, consider either combining multiple simple small questions into one question or breaking down a straightforward engineering/mathematical question into multiple connected steps, depending on the subject. This can be better explained with examples. In the former case, as demonstrated in Example-Question #4, students are given a Design of Experiments question (Latin Square Design) where they should calculate the Sum of Squares (SS) for all factors at different levels. Calculating SS for a single factor, single level is straightforward. However, when the question doesn't explicitly ask students to calculate SS for all three factors and SS for the natural error, the answer provided by ChatGPT may be inaccurate. In the latter case, let's consider a manufacturing engineering question that asks students to calculate the machine tool's life using Tyler's Tool Life Equation. Instead of providing the equation and all parameter values, the question can be divided into several parts. For example, it could involve finding key parameter values from the manufacturing/material handbook and then using those values to calculate the tool life based on the type of tools, material type, cutting speed, rpms, etc.

Guideline #5: Design questions by asking students to provide local examples or personal experiences. This approach is particularly suitable for non-STEM courses, such as social science courses. As demonstrated in Example Question #5, one of the main reasons why the instructor considers the answer provided by ChatGPT is the lack of local (in this case, a Taiwanese example) and personal observations. While AI tools can conveniently explore the question statements and prepare appropriate answers (as seen in the 6-step explanation for Question #5), they cannot link their knowledge base or training materials to local examples or incorporate detailed personal observations. It's worth noting that some Taiwanese educational institutions have adopted the practice of requiring students to disclose whether they utilized ChatGPT when preparing answers for examinations or assignments.

Guideline #6: If you are not sure whether a question can be easily solved by ChatGPT or other AI tools, simply play the questions with the ChatGPT before given to students.

IV. DISCUSSIONS AND CONCLUSIONS

In this article, we have demonstrated several guidelines that course instructors can implement when preparing online exam questions or assignment questions. However, it should be noted that these guidelines may not be as effective when dealing with reports, theses, and essays. The main advantage of AI tools like Chat-GPT is the tremendous amount of training materials they possess (for Chat-GPT 3.5, more than 175 billion training points). While these tools may not be best suited for tasks requiring specific answers or calculated values, the industrial and business communities have already shown many successful cases where tools like Chat-GPT can write poems, essays, novels, develop programming codes, generate advertising documents, and more. These results often lack a clear right/wrong distinction. Similar to earlier AI tools, such as Artificial Neural Networks (ANNs), natural language transformers like Chat-GPT operate like black boxes, making it challenging to trace how answers are generated and to guarantee their repeatability. One of the main issues with these AI tools is that they treat all training materials as 'TRUE FACTS'. Thus, if thousands of training materials erroneously state that January 1st is the Chinese New Year, this incorrect answer may be presented when prompted with the question about 'When is the Chinese New Year?'.

Nevertheless, we cannot assume that these AI tools will never overcome the barriers we place within assessment questions. Depending on the speed of development and the emergence of 'customized' AI tools in specific fields, university professors must continually prepare and update themselves regarding the current status of these issues and challenges. Fortunately, academic communities worldwide are highly vigilant about this threat and actively engaged in preparing proper assessment tools.

On a positive note, university teaching faculties can also benefit from Chat-GPT when preparing exam questions. By uploading lecture contents to Chat-GPT, it can prioritize defined domain knowledge and generate questions for quizzes and tests [10]. While the authors have not engaged in this experimental activity, there are already many reports available on this subject.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

This work was jointly prepared by the first and the second author. Specifically, M. Wang prepared the test questions from the engineering group and M. Lee prepared the question from the social science group. Both authors have approved the final version.

REFERENCES

[1] A. Bahrini, *et al*., "ChatGPT: Applications, opportunities, and threats," in *Proc. 2023 IEEE Systems and Information Engineering Design Symposium*, 2023, pp. 274–279.

[2] A. Hiken, "Marketers using AI are relying on social media for ChatGPT and DALL-E knowledge: Social platforms offer loads of AI expertise, from which brands and agencies are learning the essentials for their activations," *Advertising Age*, vol. 94, no. 5, p. 14, 2023.

[3] M. Neumann, *et. al*., "We need to talk about ChatGPT: The future of AI and higher education," in *Proc. IEEE/ACM 5th International*

Workshop on Software Engineering Education for the Next Generation, 2023, pp. 29–32.

[4] S. Nikolic, *et al.*, "ChatGPT versus engineering education assessment: A multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity," *Journal of Engineering Education*, vol. 48, pp. 559–614, 2023.

[5] E. L. Ouh, *et al.*, "ChatGPT, can you generate solutions for my coding exercises? An evaluation on its effectiveness in an undergraduate Java programming course," in *Proc. Annual Conference on Innovation and Technology in Computer Science Education*, 2023, vol. 1, pp. 54–60.

[6] P. Rajabi, *et al.*, "Exploring ChatGPT's impact on post-secondary education: A qualitative study," in *Proc. the 25th Western Canadian Conference on Computing Education*, 2023.

[7] G. Sloane, "How Meta, Google and Snap are embracing generative AI in advertising and beyond: Following the popularity of ChatGPT, the top tech companies outline their work with advanced creative tools," *Advertising Age*, vol. 94, no. 2, pp. 18–18, 2023.

[8] A. Smolansky, *et al.*, "Educator and student perspective on the impact of generative AI on assessments in higher education," in *Proc. the 10th ACM Conference on Learning*, 2023, pp. 378–382.

[9] R. P. Uhlig, "Students use of artificial intelligence to write technical engineering papers – Cheating or a tool to augment learning," in *Proc. ASEE Annual Conference and Exposition*, 2023.

[10] T. Driscoll, "ChatGPT teacher tips part 2: Quizzes and tests," *Edtechteacher Newsletter*, 2023.