EVALUATING POPULAR MOOC PLATFORMS BY GENERATIVE ARTIFICIAL INTELLIGENCE (AI) ROBOTS: HOW CONSISTENT ARE THE ROBOTS?

Victor K. Y. Chan
Faculty of Business, Macao Polytechnic University
Rua de Luis Gonzaga Gomes, Macao, China

ABSTRACT

This article intends to investigate the consistency between a few popular generative AI robots in the evaluation of massive open online course (MOOC) platforms. The four robots experimented with in the study were Claude+, GPT-4, Sage, and Dragonfly, which were tasked with awarding rating scores to the eight major dimensions, namely (1) content/course quality, (2) pedagogical design, (3) learner support, (4) technology infrastructure, (5) social interaction, (6) learner engagement, (7) instructor support, and (8) cost-effectiveness, of the 31 currently very popular MOOC platforms. Only Claude+'s and Dragonfly's rating scores turned out to be amenable to statistical analysis. For each of the two robots, the minimum, the maximum, the range, and the standard deviation of the rating scores for each of the eight dimensions were computed across all the 31 MOOC platforms. The rating score difference for each of the eight dimensions between the two robots was calculated for each platform. The mean of the absolute value, the minimum, the maximum, the range, and the standard deviation of the differences for each dimension between the two robots were calculated across all platforms. A paired sample *t*-test was then applied to each dimension for the rating score difference between the two robots over all the platforms. Finally, a correlation coefficient of the rating scores was computed for each of the eight dimensions between the two robots across all the MOOC platforms. The computational results were to reveal whether the two robots awarded discrimination in evaluating each dimension across the platforms, whether any of the two robots systematically underrated or overrated any dimension with respect to the other robot, and whether there was consistency between the two robots in evaluating each dimension across the platforms. It was found that discrimination was prominent in the evaluation of all dimensions save Dragonfly's rating of the dimensions learner support, technology infrastructure, and instructor support, Claude+ systematically underrated all dimensions ($p < 0.000 < 0.05$) compared with Dragonfly except for the dimension cost-effectiveness, which Claude+ systematically overrated ($p = 0.003 < 0.05$), and the evaluation by the two robots was consistent only for the dimensions content/course quality, pedagogical design, and learner engagement with the correlation coefficients ranging from 0.445 to 0.632 (p from 0.000 to 0.012 < 0.05). Consistency implies at least the partial trustworthiness of the evaluation of these MOOC platforms by either of these two popular generative AI robots based on the analogous concept of convergent validity for an operationalized instrument to measure an abstract construct.

KEYWORDS

Massive Open Online Course Platforms, MOOC Platforms, Artificial Intelligence, Evaluation, Consistency

1. INTRODUCTION

Massive open online courses (MOOCs) have become a popular mode of learning in the last decade. The widespread adoption of MOOCs has led to the creation of different MOOC platforms with varying features and capabilities. Evaluating such features and capabilities of a MOOC system is essential to ensure that it meets the needs of learners and instructors. Such traditional means of evaluation as surveys and user feedback, are time-consuming and subjective and can usually cover a limited sample of respondents' opinions. The advent of generative artificial intelligence (AI) robots may potentially offer a nascent solution to this evaluation difficulty such that MOOC platforms can be alternatively evaluated and compared by robots in an automated manner. Having said this, there exists no absolutely standard "baseline" against which a particular evaluation modality can be benchmarked in order to certify the evaluation's trustworthiness. All one can do to ascertain which evaluation to trust is to gauge the consistency between multiple evaluations. If all evaluations turn out to be consistent to an extent, the odds are that all of them are trustworthy, although theoretically all being erroneous can never be ruled out. This is analogous to the concept of convergent validity for an operationalized instrument to measure an abstract construct. This article attempts to explore such consistency between a few popular generative AI robots in the evaluation of MOOC platforms.

1.1 Generative AI Robots for Evaluating and Comparing MOOC Platforms

Generative AI robots, also known as generative models or generative adversarial networks (GANs), are a type of artificial intelligence that can create new and original content such as images, music, or text. These robots are trained on large datasets and use complex algorithms to generate content that is similar to the training data but also unique in its own way. One example of a generative AI robot is OpenAI's ChatGPT language model, which is capable of generating human-like text with a high degree of coherence and creativity. (Wang et al., 2017)

For the evaluation of MOOC platforms, generative AI robots can be programmed or otherwise instructed to analyze the platforms' such perspectives or dimensions as (1) content/course quality, (2) pedagogical design, (3) learner support, (4) technology infrastructure, (5) social interaction, (6) learner engagement, (7) instructor support, and (8) cost-effectiveness, among many others. Thereby, each robot generates an objective and standardized rating score for each dimension of each platform, serving the purpose of evaluation.

In recent years, researches have studied the application of generative AI robots to MOOC-related phenomena. One monumental one conducted by Brinton et al. (2014) developed a unified generative model algorithm for the discussion threads, which allowed the researchers both to choose efficient thread classifiers and to design an effective algorithm for ranking users' thread relevance in the courses offered by a major MOOC provider. Their algorithm was compared against two baselines using human evaluation from Amazon Mechanical Turk.

Another study by Li and Xing (2021) examined the extent to which the deep-learning-based natural language generation (NLG) models can offer responses similar to human-generated responses to the learners in MOOC forums. Specifically, under the framework of social support theory, this study examined the use of then state-of-the-art deep learning models recurrent neural network (RNN) and generative pre-trained transformer 2 (GPT-2) to provide students with informational, emotional, and community support with NLG on discussion forums. The results showed that GPT-2 outperformed RNN on all measures. The results showed GPT-2 model could provide supportive and contextual replies to a similar extent compared to humans.

In contrast, the author is not aware of any extant literature specifically evaluating MOOC platforms by means of generative AI robots. This is exactly the gap that this article is to fill.

2. METHODOLOGY

2.1 Data and Materials

The present study experimented with four very popular generative AI robots, namely Claude+, Dragonfly (de Souza et al., 2023), GPT-4 (Zhang et al., 2023), and Sage (de Souza et al., 2023) as candidates for the evaluation and comparison, all of them being available through the AI portal poe.com. Eight major dimensions to evaluate a MOOC platform were identified (Albelbisi, 2020; Hew and Cheung, 2014; Khalil and Ebner, 2014; Kizilcec, Piech, and Schneider, 2013; Liyanagunawardena, Adams, and Williams, 2013) as (1) content/course quality, (2) pedagogical design, (3) learner support, (4) technology infrastructure, (5) social interaction, (6) learner engagement, (7) instructor support, and (8) cost-effectiveness, which were to be rated by the robots in this study. Content/course quality measures the overall quality and relevance of the course content, including the course design, instructional strategies, and assessment methods. It is essential to ensure that the course content is up-to-date, accurate, and relevant to the learners' needs. The quality of the MOOC content is a critical factor that affects learners' satisfaction with the course. Pedagogical design refers to the design of the course, including the teaching methods, assessment strategies, and learning outcomes. It is essential to ensure that the course is designed in a way that encourages active learning and promotes learner engagement. The pedagogical design of a MOOC is a critical factor that affects learners' engagement and motivation. Learner support includes the support provided to learners throughout the course. It is essential to ensure that learners have access to adequate support, including technical support and academic support. Learner support is a critical factor that affects learners' completion rates and overall satisfaction with the course. Technology infrastructure measures the technological capabilities of the MOOC platform, including its ability to deliver course content, interact with learners, and manage learner data as well as how user-friendly, accessible, and reliable it is. Technology infrastructure may be assessed by metrics like platform uptime, speed of content delivery, compatibility with different devices and browsers, and security measures implemented by the platform. Social interaction refers to the opportunities for learners to interact with each other and with the instructors. It is essential to ensure that learners have opportunities to collaborate, discuss, and share ideas with each other. Social interaction is a critical factor that affects learners' engagement and satisfaction with the course. Learner engagement measures the level of interaction between learners and the course content. Engagement is a crucial factor in determining the effectiveness of a MOOC system as it affects the learning outcomes of learners. It can be appraised by such metrics as the average time spent on the course content, the number of interactions with the course materials, the number of forum posts and comments by learners, and the average completion rate of the course. Instructor support measures the level of support provided to learners by the instructors. Instructor support is important because it fosters a sense of community and increases learner engagement. It can be gauged by metrics such as the response time to learner queries, the quality of responses to learner queries, the availability of instructors during course hours, and the frequency of instructor-led sessions. Cost-effectiveness measures the cost of delivering the course content and the benefits derived from it. Cost-effectiveness is important because it determines the viability of a MOOC system as a mode of delivering education. It can be translated into metrics, namely, the cost per learner, the return on investment of the MOOC system, the cost savings compared to traditional modes of education delivery, and the revenue generated by the MOOC system.

Then, the data collection was kicked off by submitting the following request to both Sage and Dragonfly:

> "Please name 20 most popular MOOC platforms."

Sage replied with the following list in alphabetical order:

> "1. Alison 2. Canvas Network 3. Coursera 4. edX 5. Eliademy 6. FutureLearn 7. Google Classroom 8. Harvard Online Learning 9. Iversity 10. LinkedIn Learning 11. Lynda.com 12. Miríada X 13. NovoEd 14. Open2Study 15. Pluralsight 16. Skillshare 17. Udacity 18. Udemy 19. UNED COMA 20. XuetangX,"

whereas Dragonfly replied with a slightly different list:

> "1. Coursera 2. edX 3. Udacity 4. Udemy 5. Khan Academy 6. Alison 7. FutureLearn 8. Open2Study 9. Saylor Academy 10. Canvas Network 11. OpenLearning 12. Academic Earth 13. Open Education by Blackboard 14. Open Yale Courses 15. MIT OpenCourseWare 16. Carnegie Mellon Open Learning Initiative 17. Harvard Extension School 18. Stanford Online 19. Rice University OpenStax 20. Google Digital Garage"

Eradicating 8 overlapping platforms, an ultimate list of 32 most popular MOOC platforms was compiled.

In other words, this study empowered Sage and Dragonfly to determine which MOOC platforms were most popular and were to be covered by the subsequent procedures. Then, the following request was submitted to Claude+, Dragonfly, GPT-4, and Sage:

> "For each of the eight dimensions (1) content/course quality, (2) pedagogical design, (3) learner support, (4) technology infrastructure, (5) social interaction, (6) learner engagement, (7) instructor support, and (8) cost-effectiveness, please give a rating score to each of the popular MOOC platforms (namely, Coursera, edX, Udacity, Udemy, Khan Academy, Alison, FutureLearn, Open2Study, Saylor Academy, Canvas Network, OpenLearning, Academic Earth, Open Education by Blackboard, Open Yale Courses, MIT OpenCourseWare, Carnegie Mellon Open Learning Initiative, Harvard Extension School, Stanford Online, Rice University OpenStax, Google Digital Garage, Eliademy, Google Classroom, Harvard Online Learning, Iversity, LinkedIn Learning, Lynda.com, Miríada X, NovoEd, Pluralsight, Skillshare, UNED COMA, and XuetangX or as large a subset of them as you like) based on a scale of 1 to 10 (1 being the worst and 10 the best). Please derive your scores from global users' textual comments on these eight dimensions of these platforms as appear all around the web."

It is noteworthy that the above request explicitly spelt out the 32 MOOC platforms on the ultimate list.

Both Claude+ and Dragonfly replied with complete rating scores for all the eight dimensions and all the 32 platforms except that the former only rated five out of the eight dimensions for the platform Google Digital Garage, leaving only 31 platforms for subsequent analysis. In contrast, GPT-4 only rated the eight dimensions solely for five platforms Coursera, edX, Udacity, Udemy, and Khan Academy, which were quite insufficient for further analysis, whilst Sage rated the eight dimensions merely for the two platforms Coursera and edX and rated some of the eight dimensions for another seven platforms Khan Academy, FutureLearn, OpenLearning, MIT OpenCourseWare, Harvard Extension School, Stanford Online, and Harvard Online Learning. The implication was that only Claude+'s and Sage's rating scores for 31 platforms (i.e., the 32 platforms on the ultimate list save Google Digital Garage) were amenable to further analysis, whereas those of GPT-4 and Sage were too fragmented for that purpose. Please note that both the request above expressly pinpoint "…derive your scores from global users' textual comments on these eight dimensions of these platforms as appear all around the web." Stated differently, each robot presumably derived its rating scores from global users' textual comments appearing all on the worldwide web instead of parroting any comparable scores already existing somewhere.

2.2 Analysis

For each of the two robots, the minimum, the maximum, the range, and the standard deviation of the rating scores for each of the eight dimensions were computed across all the 31 MOOC platforms (i.e., 32 platforms on the ultimate list excluding Google Digital Garage). If there is an appreciable range and standard deviation for a particular dimension, it is implied that the robot concerned awards discrimination in rating the dimension across the platforms.

Then, the rating score difference for each of the eight dimensions between the two robots was calculated for each of the 31 platforms. The mean of the absolute values, the minimum, the maximum, the range, and the standard deviation of the differences for each dimension between the two robots were calculated across all the 31 platforms. If the mean of the absolute values, the range, and the standard deviation for a particular dimension are sufficiently small, it is uncovered that the two robots neither overrate nor underrate erratically with respect to each other the dimension across the platforms. A paired sample *t*-test was then applied to each dimension for the rating score differences between the two robots over all the 31 platforms. If the *t*-test is significant for a particular dimension and the corresponding mean difference is positive (negative), it is revealed that the first robot systematically overrates (underrates) the dimension with respect to the second robot.

Finally, a correlation coefficient of the rating scores was computed for each of the eight dimensions between the two robots across the 31 MOOC platforms. If the correlation coefficient is positively high (for instance, over 0.6) for a particular dimension, it is indicated that there is consistency between the two robots in rating the dimension across the platforms.

3. RESULTS

Table 1 lists the minimum, the maximum, the range, and the standard deviation of the rating scores for each of the eight dimensions across the 31 MOOC platforms as rated by each of the two robots Claude+ and Dragonfly. Whilst Claude+ rated all the eight dimensions with substantial discrimination, Dragonfly rated with sufficient discrimination only the five dimensions content/course quality, pedagogical design, social interaction, learner engagement, and cost-effectiveness but not the remaining three dimensions learner support, technology infrastructure, and instructor support as manifested by zero ranges and standard deviations. More specifically, Claude+ also assigned exceptionally strong and weak discriminations to the dimensions pedagogical design and cost-effectiveness, respectively, as per their corresponding larger and smaller standard deviations relative to those of other dimensions.

Table 1. The minimum, the maximum, the range, and the standard deviation of the rating scores for each of the eight dimensions across the 31 MOOC platforms as rated by each of the two robots Claude+ and Dragonfly

Robot (sample size *n*)	Minimum/ maximum/ range/ standard deviation	Content/ course quality	Pedagogical design	Learner support	Technology infrastructure	Social interaction	Learner engagement	Instructor support	Cost-effectiveness
Claude+ (*n* = 31)	Minimum	3	2	1	2	1	2	1	7
	Maximum	8	9	7	8	8	9	8	10
	Range	5	7	6	6	7	7	7	3
	Standard derivation	1.4876	2.1241	1.4760	1.6688	1.6707	1.7339	1.5658	0.9298
Dragonfly (*n* = 31)	Minimum	8	8	8	8	7	8	8	7
	Maximum	9	9	8	8	8	9	8	9
	Range	1	1	0	0	1	1	0	2
	Standard derivation	0.4448	0.4448	0.0000	0.0000	0.3005	0.4448	0.0000	0.4973

Table 2 enumerates the mean of the absolute values, the minimum, the maximum, the range, and the standard deviation of the rating score differences for each of the eight dimensions across the 31 MOOC platforms between the two robots. There seems to be more discrepancy between Claude+ and Dragonfly in rating the seven dimensions content/course quality, pedagogical design, learner support, technology infrastructure, social interaction, learner engagement, and instructor support than in rating the remaining dimension cost-effectiveness in view of the means of the absolute values, the ranges, and the standard deviations of the corresponding rating score differences for the former seven dimensions being far greater than those for the latter dimension. In other words, the two robots overrate or underrate erratically with respect to each other the former seven dimensions across the platforms more than the latter dimension.

Table 2. The mean of the absolute values, the minimum, the maximum, the range, and the standard deviation of the rating score differences for each dimension between the two robots

Differences (sample size *n*)	Mean of the absolute values/ minimum/ maximum/ range/ standard deviation of the differences	Content/ course quality	Pedagogical design	Learner support	Technology infrastructure	Social interaction	Learner engagement	Instructor support	Cost-effectiveness
Claude+ – Dragonfly (*n* = 31)	Mean of the absolute values	2.5484	2.8710	4.6129	2.4194	4.3871	3.4194	4.5806	0.7742
	Minimum	-5	-6	-7	-6	-7	-6	-7	-1
	Maximum	0	0	-1	0	0	0	0	2
	Range	5	6	6	6	7	6	7	3
	Standard derivation	1.3376	1.8751	1.4760	1.6688	1.7828	1.5869	1.5658	0.8896

Table 3 illustrates the paired sample *t*-tests of the rating score differences for each of the eight dimensions between the two robots over the 31 MOOC platforms. With respect to Dragonfly, Claude+ appears to underrate the seven dimensions content/course quality, pedagogical design, learner support, technology infrastructure, social interaction, learner engagement, and instructor support and overrate the remaining dimension cost-effectiveness at the 1% significance level ($p < 0.01$).

Table 3. The paired sample *t*-test of the rating score differences for each of the eight dimensions between the two robots

Differences (sample size *n*)	Dimension	Mean difference / [95% confidence interval]	*t* (*p*-value) / degrees of freedom
Claude+ – Dragonfly (n = 31)	Content/course quality	-2.548 / [-3.039, -2.058]	-10.607 (0.000**) / 30
	Pedagogical design	-2.871 / [-3.559, -2.183]	-8.525 (0.000**) / 30
	Learner support	-4.613 / [-5.154, -4.072]	-17.401 (0.000**) / 30
	Technology infrastructure	-2.419 / [-3.031, -1.807]	-8.072 (0.000**) / 30
	Social interaction	-4.387 / [-5.041, -3.733]	-13.701 (0.000**) / 30
	Learner engagement	-3.419 / [-4.001, -2.837]	-11.997 (0.000**) / 30
	Instructor support	-4.581 / [-5.155, -4.006]	-16.289 (0.000**) / 30
	Cost-effectiveness	0.516 / [0.190, 0.842]	3.230 (0.003**) / 30

** $p < 0.01$

Table 4 depicts the correlation coefficient of the rating scores for each of the eight dimensions between the two robots over the 31 platforms, the 95% conference interval for the correlation coefficient, and the *p*-value to test whether the coefficient differs from zero. The two robots are highly, positively correlated and thus consistent in rating the three dimensions content/course quality, pedagogical design, and learner engagement as attested by their corresponding positive correlation coefficients and by the corresponding *p*-values being less than either 0.01 ($p < 0.01$) or 0.05 ($p < 0.05$). It is worth noting that a highly positive correlation coefficient and thus consistency for a particular dimension imply a high rating of a platform for the dimension by one robot being generally associated with a high rating of that platform for that dimension by another robot and vice versa albeit these two ratings may not necessarily be the same or not even close. There is no evidence of consistency for other dimensions.

Table 4. The correlation coefficient of the rating scores for each of the eight dimensions between the two robots

Dimension	Correlation coefficient / [95% confidence interval]	*p*-value
Content/course quality	0.470 / [0.1388, 0.7067]	0.008**
Pedagogical design	0.632 / [0.3578, 0.8059]	0.000**
Learner support	-[a]	-[a]
Technology infrastructure	-[a]	-[a]
Social interaction	-0.296 / [-0.5886, 0.06518]	0.107
Learner engagement	0.445 / [0.1076, 0.6905]	0.012*
Instructor support	-[a]	-[a]
Cost-effectiveness	0.346 / [-0.009506, 0.6239]	0.056

* $p < 0.05$
** $p < 0.01$
[a] The rating scores by at least one robot are constant for the dimension over all the 31 platforms, rendering no correlation coefficient.

In summary, whereas it may be rather safe to trust and rely on Claude+'s and Dragonfly's ratings of MOOC platforms for such dimensions as content/course quality, pedagogical design, and learner engagement, one is better off distrusting those for the remaining dimensions learner support, technology infrastructure, social interaction, instructor support, and cost-effectiveness.

4. CONCLUSION

There are quite some factors underlying inconsistency between generative AI robots in the evaluation of MOOC platforms (or, in fact, anything). As detailed above, the author also submitted to another generative AI robot Sage on poe.com a request resembling the one submitted to Claude+ and Dragonfly, yielding the following respective insightful preamble and epilogue in the robot's reply:

"As an AI language model, I can provide you with general information on the MOOC platforms you mentioned, but it's important to note that deriving scores based purely on global user comments can be subjective and may not accurately reflect the quality of these platforms. Additionally, there may be bias or variation in user experiences based on factors such as their background, course selection, and level of engagement."

"It's important to note that these ratings are based on general user comments and experiences, and may not accurately reflect the experience of every individual user. Additionally, the ratings can vary depending on the specific course or program taken on each platform."

Whereas generative AI robots are valuable in analyzing global users' textual comments at scale and offering high-level insights into the strengths, weaknesses, and reputations of various MOOC platforms by rating each platform in omnibus dimensions using some precise, standardized scores, there are a series of shortcomings inherent in such robots. Such shortcomings and thus factors underlying inconsistency between different robots can be summarized with the points below.

1. Textual user comments are contingent upon the courses and programs concerned. Even for the same platform, user comments may be disparate due to the different courses and programs that the users study.
2. Textual user comments are so subjective and prone to bias or variation that even for the same platform, user comments may vary drastically in accordance with the particular users involved.
3. Out of the dramatically varying pool of user comments, a particular robot's ratings of a particular platform are thus very much dependent on the sample of user comments extracted from the pool for the robot's training. Therefore, it is not uncommon to see two or more robots' ratings of the same platform differ as a result of their different training samples of user comments.

Aside from inconsistency, generative AI robots' evaluation and comparison of MOOC platforms are subject to other limitations. One of them is the reliance on particular robots' pre-defined rules and parameters, which may not be able to capture the full range of features and capabilities of some platforms.

Despite probable inconsistency and other limitations, generative AI robots admittedly have emerged as at least a promising alternative method for objectively evaluating and comparing MOOC platforms. By serving as a standardized and objective method of evaluation, generative AI robots can assist educators, learners, educational institutions, and decision-makers in choosing the most suitable MOOC platform for their needs.

This study itself is not without its critics. First, only two generative AI robots Claude+ and Dragonfly managed to generate data compendious enough for analysis in this study. Second, these two robots were trained on data up to a few years back, so the ratings were by no means reflective of the "most" current and latest MOOC platform versions.

Future research can derive ways to avert the above limitations and enhance the use of generative AI robots for evaluating and comparing MOOC platforms. In addition, this study's coverage and currency of the two generative AI robots can be extended and updated to more robots trained on more recent data, in particular, upon the expected advent of new generative AI robots in the near future. Moreover, generative AI robots can be in collaboration with traditional methods, such as surveys, user feedback, and statistical analysis, to assess MOOC platforms more holistically.

REFERENCES

Albelbisi, N. A., 2020. Development and Validation of the MOOC Success Scale (MOOC-SS). *In Education and Information Technologies*, Vol. 25, No. 5, pp. 4535-4555.

Brinton, C. G et al, 2014. Learning about Social Learning in MOOCs: From Statistical Analysis to Generative Model. *In IEEE Transactions on Learning Technologies*, Vol. 7, No. 4, pp. 346-359. doi: 10.1109/TLT.2014.2337900.

de Souza, C. et al, 2023. Are the New AIs Smart Enough to Steal Your Job? IQ Scores for ChatGPT, Microsoft Bing, Google Bard and Quora Poe. Retrieved May 10, 2023 from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4412505

Hew, K. F. and Cheung, W. S., 2014. Students' and Instructors' Use of Massive Open Online Courses (MOOCs): Motivations and Challenges. *In Educational Research Review*, Vol. 12, pp. 45-58. doi: 10.1016/j.edurev.2014.05.001

Khalil, H. and Ebner, M., 2014. MOOCs Completion Rates and Possible Methods to Improve Retention - A Literature Review. *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications*. pp. 1236-1244.

Kizilcec, R. F. et al, 2013. Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses. *Proceedings of the Third ACM International Conference on Learning Analytics and Kwonledge*. pp. 170-179. doi: 10.1145/2460296.2460330

Li, C. and Xing, W., 2021. Natural Language Generation Using Deep Learning to Support MOOC Learners. *In International Journal of Artificial Intelligence in Education*, Vol. 31, No. 2, pp. 186-214. doi: 10.1007/s40593-020-00235-x

Liyanagunawardena, T. R. et al, 2013. MOOCs: A Systematic Study of the Published Literature 2008-2012. *In The International Review of Research in Open and Distributed Learning*, Vol. 14, No. 3, pp. 202-227. doi: 10.19173/irrodl.v14i3.1455

Wang, K. et al, 2017. Generative Adversarial Networks: Introduction and Outlook. *In IEEE/CAA Journal of Automatica Sinica*, Vol. 4, No. 4, pp. 588-598. doi: 10.1109/JAS.2017.7510583.

Zhang, C. et al, 2023. A Complete Survey on Generative AI (AIGC): Is ChatGPT from GPT-4 to GPT-5 All You Need? arXiv:2303.11717v1 [cs.AI] Retrieved May 10, 2023 from https://arxiv.org/abs/2303.11717