



# Language Models: Viable Strategies for Portfolio Assessment

Per Arne Godejord

Nord University Business School, Norway

## Abstract

*In the wake of heightened speculation surrounding ChatGPT, a prominent language model, this paper delves into the potential implications for portfolio assessment within university courses. ChatGPT, employing statistical computations to generate written text, has raised concerns about its impact on exam submissions conducted outside conventional settings [1]. This study, conducted from December 2022 to December 2023, assesses the viability of ChatGPT and similar language models in the context of portfolio assessment for university courses, specifically within ICT and Learning 1, ICT and Learning 2, and the MBA course ORG5005 – Digital Preparedness.*

*Seven language model systems underwent evaluation based on their ability to generate proficient academic responses, emphasizing critical thinking and practical application. Tasks spanned information security, copyright matters, project work in ICT education, and digital preparedness. Despite the systems' endeavours, the results revealed resolute failure in addressing the specified tasks. Students relying solely on ChatGPT or similar systems without scholarly exploration and source analysis would receive an overall failing grade, lacking academic reflection and practical solutions.*

*While the assessed systems fell short in producing satisfactory academic responses, this study refrains from dismissing the potential for future advancements. The failure of the current models does not negate the possibility that continued progress in ChatGPT, and comparable systems may address the extensive work requirements characteristic of the ICT and Learning study programmes and ORG5005 – Digital Preparedness in subsequent developments.*

**Keywords:** *Portfolio Assessment, Continuing education, Language Models*

## 1. Introduction

Language models, such as ChatGPT, Jenny, and Bing Chat, have garnered attention for their potential applications in assisting students with academic assignments. This paper describes a set of tests of the performance of these language models in completing university assignments across a set of courses offered by Nord University. By examining their ability to generate academic texts and reflection notes, the tests were seeking to assess the ability of language models to actually produce high level academic text in response to a set of assignments. This paper describes the results of one year of testing conducted in the period December 2022 to December 2023.

## 2. Background

Nord University offers a diverse range of online courses specifically designed for professionals in full-time employment. These courses emphasize the integration of practical and theoretical knowledge, using portfolio assessments as a common evaluation method. The assignments are kept as authentic as possible, to enable students to combine their professional everyday practice with both practical and theoretical work. This style of assignments enables students to demonstrate their individual skills and reflect on their learning experiences.

### 2.1 Selected Courses for Study

The selected courses for this study include ICT and learning programs tailored for professional teachers and instructors, as well as one MBA course "ORG5005-Digital Preparedness," which focuses on digital security and preparedness tailored for professional officers of the armed forces, police, and paramedics, as well as other officials involved in various preparedness work. These courses were chosen because they are 100% online and feature assignments constructed according to Bloom's Taxonomy, emphasizing both practical and theoretical work. This necessitates students combining theoretical knowledge with personal workplace experience. Additionally, all these courses employ



portfolio assessment as their examination system. This approach allows students to display practical and theoretical skills while reflecting on their learning experiences.

From the ICT and learning 1 and 2 study programs, the following three courses were selected:

#### *ICT1013-Basic Information Security*

This introductory course covers fundamental terms and concepts related to information security. Its objective is to enhance students' awareness of their responsibility for digital preparedness, both as private individuals and employees. The course content includes an elementary understanding of various cyber threats.

#### *ICT1016-Digital Literacy*

This course focuses on concepts related to digital skills, digital competence, and digital education. It places significant emphasis on copyright issues and ethical guidelines for using digital media in educational and instructional contexts.

#### *ICT1024-Project Work with ICT*

This course examines the use of various digital tools in project work and how virtual teams can collaborate on projects from the initial idea through preliminary development to project implementation, using various cloud-based support tools.

From the MBA-programme "Preparedness and Emergency Management" the following course was chosen:

#### *ORG5005-Digital Preparedness*

The "ORG5005-Digital Preparedness" course provides an in-depth understanding of current digital threats and vulnerabilities, along with relevant protective measures for daily preparedness and high-level crisis scenarios. The course aims to foster reflective thinking in students to enhance their emergency management skills and overall security awareness as citizens of a digital society. It does not train students to combat hackers directly but focuses on understanding the role of personal attitudes and digital preparedness in improving the overall readiness of their employers and the nation. Additionally, students should be capable of conducting awareness-raising activities in their organizations and participating in preparedness exercises that address the loss of digital infrastructure.

## **2.2 Course Structure and Teaching Methodology**

All activities within these courses are asynchronous, and teaching consists of student-active assignments that combine practical and theoretical work. The students must link both the practical as well as the theoretical work to their own professional everyday life and their individual experiences as teacher, instructors, course managers or army officer.

A critical component of the theoretical parts of the assignments is the "Reflection Notes" chapter, where students demonstrate their ability to reflect on their own learning processes related to the assignments. This reflective practice is essential for deepening their understanding and integration of practical and theoretical knowledge.

None of the courses have traditional exams but rely on portfolio assessments. Students must complete two obligatory assignments, submit drafts for evaluation and guidance, and then submit their final versions. All assignments combine practical and theoretical tasks, requiring students to relate theory to

their individual work life experiences. This alignment with the upper levels of Bloom's Taxonomy ensures a comprehensive evaluation of student competencies.

Portfolio assessment in these courses complies with what is known as "instructional" or "working" portfolios, allowing students to demonstrate both practical and theoretical skills. Students cannot choose which tasks to submit for final evaluation but must submit all required tasks.



In summary, Nord University's online courses are meticulously designed to integrate practical experience with theoretical knowledge through portfolio assessment. This ensures that professionals in full-time employment can enhance their skills and reflect critically on the various subjects of the assignments and their own learning process, thereby contributing to their personal and professional development. The big question always arising concerning evaluation in courses without traditional exams, is who is really writing the assignments? And more pertinently since November of 2022, are the assignments written by ChatGPT or similar language models?

### 3. The Tests

To evaluate how resilient my assignments are to the use of various language models, such as ChatGPT and Bing Chat (now Microsoft copilot), I submitted the complete text of various assignments to the following seven systems:

- ChatGPT (OpenAI, free version)
- GPT UiO (University of Oslo Norwegian version based on ChatGPT 3.5 turbo)
- Sikt KI-Chat (Norwegian Agency for Shared Services in Education and Research version based on ChatGPT 4)
- GPT-3 Playground (OpenAI, Free version)
- Chatsonic (WriteSonic, Free version)
- Bing Chat (Ordinary version, part of Edge Browser)
- Jenni (Free version)

I then asked the systems to produce academic texts to the various assignments. The systems were also asked to write a reflection note based on one of the assignments presented.

### 4. Results

Results indicate significant shortcomings in the responses generated by the various language models. While all the chosen systems demonstrate factual accuracy on par with the Lower order thinking skills of Blooms Taxonomy, like repeating simple facts found online or in their databases, they lacked analytical depth and fail to meet the academic criteria for higher order thinking. In the following I will briefly describe the results, and I start with the MBA course.

#### 4.1 The course "*ORG5005-Digital Preparedness*"

The assignments given in this course were both formulated as a 5 paragraph order (SMEAC), and the first assignment focused on the student's ability to create an exercise directive for a possible exercise for the students own organisation, with a focus on hybrid attacks, where the loss/disruptions of digital systems has serious consequences for their organisations ability to perform their work. The exercise directive had to have a comment note attached explaining and justifying every choice made in constructing the directive, as well as a reflection note describing the students individual learning process,

The response from all the systems evaluated was to repeat the text of the assignment with slightly altered wording and the result was naturally nowhere near fulfilling the demands of the assignment.

#### 4.2 The course "*ICT1013-Basic Information Security*"

The systems were given the first task in this course and asked to provide an academic text describing the concept of Social Engineering, and then analysing in what way data from mobile devices can help strengthening this form of attack and how their own security attitudes and use of digital media could influence on the digital security of their own work place. The assignment pointed to both one of my video lectures and two relevant articles available in the Canvas course room. The text produced fell significantly short of the requirements in the field of ICT and learning studies. All the system's reviews of "Social engineering" were superficial and failed to meet even the basic criteria for a level 1 answer according to Bloom's taxonomy. Similarly, the analysis of mobile devices and their role in preparations for Social Engineering attacks, as well as reflection on their own attitudes and use of digital media



lacked analytical depth and professional reflection. However, it is worth noting that there were no factual errors in the answer given by the systems.

#### **4.3 The course ICT1016 Digital Literacy**

The assignment chosen from this course focused on one exception from Norwegian Copyright Law, that even if teachers may copy a certain amount from for instance a book to use in their classroom teaching, they cannot implement parts of various books, articles or media files and create a standalone work. The answers provided by the systems were all incorrect, and this highlights a significant weakness in this type of systems, its database. The language models calculated (presumably) a high statistical probability that the user is seeking information about ordinary copying based on Norwegian law and discarded any information (if such information existed in their databases) on the exemption from the general rule. Notably, Bing Chat, with its access to the internet, performed slightly better in this regard compared to the other systems.

#### **4.4 The course ICT1024 Project Work with ICT**

Here the chosen language models were presented with two assignments. The first asked the students to work in Virtual Teams and then develop an outline for a project set to use various digital media in their own classrooms to facilitate learning within a chosen school topic. The project outlines the language models produced were too simple to be submitted as an assignment paper and they were also not tailored to any students actual practical teaching reality. But the sketches presented could function as a set of key words to help students to get started with the assignment.

The language models were then given the second assignment in this topic, asking the students to generate a project report based on their practical work with their initial project outline in their classrooms. Instead of generating a report/draft report, the systems only respond with a table of contents based on the assignment text. But they did the necessary statistical calculations to connect the request to the previous assignment, and it did seem impressive. But the systems did not really do anything other than repeat the assignment text and did not present anything that could be used as a final project report as asked by the assignment.

#### **4.5 Reflections on individual learning**

In both the ICT and Learning Studies and in the MBA course ORG5005, the students must reflect on their own learning process with each assignment. This is done in a separate chapter called "Reflection note." ChatGPT and the six other systems produced coherent texts when asked, but upon closer examination it was obvious that the text was no actual reflection upon learning but a statistically generated composition of words and sentences that gave an outline that was far from what is the requirements in the assignments evaluated. For both the ICT and Learning Studies and the MBA course ORG5005, the students must reflect on their learning process based on a set of requirements found in a "Guidelines for portfolio assessment" and these guidelines are not included in the databases of the language models, nor readily available on Internet. It was obvious that language models assessed was unable to create an individual reflection on the learning process with the assignment.

### **5. Conclusion**

In conclusion, the performance of the chosen language models in completing the academic assignments given to them falls short of meeting the standards expected. While these models show potential in aiding students with simple content generation, they lack the analytical depth and critical thinking skills necessary for success in creating high level results.

The findings of the tests described in this paper are in accordance with similar test showing that language models do poorly when presented with authentic assignments, i.e. assignments intricately linked to the student's professional occupation and focusing on the individual student's ability to make judgements and decisions [2]. This underscores the importance of constructing university assignments to focus on higher order skills such as critical thinking and genuine individual reflection, qualities that language models currently lack. While these models help in simple content generation on the lowest

level, they cannot replace students' intellectual engagement and analytical reasoning [3]. Despite the hype surrounding the various language models, these systems do nothing more than performing statistical generated sentences lacking the analytical depth, critical thinking skills, and genuine reflection necessary to manage university assignments of the type described in this paper.

These findings shows that assignments demanding that students link their practical and theoretical work to their own professional everyday life and experiences, having to explain their choices and link all arguments to academical references, as well as demanding individual reflections on their own learning process, makes portfolio assessment in asynchronous online courses a viable way to assess student work even in the age of language models.

Further work will focus on doing more in depth testing to see if the ongoing developments of the various language models, enables the various systems available to the public to better manage university assignments of the type described in this paper.

## REFERENCES

- [1] Susnjak, Teo, 2022. ChatGPT: The End of Online Exam Integrity? Retrieved from: [https://www.researchgate.net/publication/366423865\\_ChatGPT\\_The\\_End\\_of\\_Online\\_Exam\\_Integrity](https://www.researchgate.net/publication/366423865_ChatGPT_The_End_of_Online_Exam_Integrity)
- [2] Lodge, Jason M., 2023. ChatGPT consistently fails (most parts of) the assessment tasks I assign my students. Here's why. Retrieved from: <https://www.linkedin.com/pulse/chatgpt-consistently-fails-most-parts-assessment-tasks-jason-m-lodge>
- [3] Mitchell, Melanie, 2023. Can Large Language Models Reason? Retrieved from: <https://aiguide.substack.com/p/can-large-language-models-reason>