

# Artificial Intelligence as a Tool to Support Task Design in the Flipped Classroom

*Pšenák Peter<sup>1</sup>, Pšenáková Ildikó<sup>2</sup>*

DOI: <https://doi.org/10.53349/re-source.2026.is1.a1543>

## Abstract

The flipped classroom model shifts the acquisition of new content to students' independent preparation, allowing contact time to be dedicated to active learning and practical application through creative tasks. The efficacy of this pedagogical approach depends on the quality of activities designed to foster the application of knowledge, skill development, and collaborative engagement. This paper examines the potential of artificial intelligence (AI), particularly generative models, as a tool for supporting the design of such tasks during the in-class phase. By exploring the utility of AI in developing problem-based scenarios, role plays, and project-based activities, the paper demonstrates how these technologies can assist educators in creating in-class experiences that are meaningful, engaging, and effective for university students.

*Keywords:* Flipped Classroom, Active Learning, Creative Task Design, AI in Education

## 1 Introduction

The higher education sector is undergoing a significant transformation, driven by both pedagogical innovation and emerging technologies. Amongst the most prominent and successful models to have emerged is the Flipped Classroom (FC). This model inverts traditional university pedagogy by shifting direct instruction and content delivery to a pre-class, independent student preparation phase, often via recorded lectures, readings, or interactive modules (Bergmann & Sams, 2012). This strategic shift is designed to liberate valuable in-class contact time, transforming it from a passive listening experience into a dynamic, active-learning workshop.

---

<sup>1</sup> Univerzita Komenského v Bratislave, Fakulta managementu, Odbojárov 10, P. O. BOX 95, 820 05 Bratislava 25, Slovakia.

*E-Mail:* [psenak9@uniba.sk](mailto:psenak9@uniba.sk)

<sup>2</sup> Trnavská univerzita v Trnave, Pedagogická fakulta, Priemyselná 4, P. O. BOX 9, 918 43 Trnava, Slovakia.

However, the success of the FC model is entirely dependent on the quality and nature of these in-class activities. Simply removing the lecture is not enough. The freed-up time must be filled with meaningful, engaging tasks that foster peer instruction, problem-solving, and the deep application of knowledge (Baig & Yadegaridehkordi, 2023). This crucial time must be filled with creative tasks that foster the application of knowledge, the development of skills, and collaboration (Pšenáková & Pšenák, 2024).

This presents a significant pedagogical bottleneck. Designing these high-impact activities, such as problem-based scenarios, complex case studies, role-plays, and collaborative projects, is a non-trivial, time-consuming work (Schallert-Vallaster et al., 2021). It requires a different skill set than creating a traditional lecture and places a heavy burden on educators.

## 1.1 Artificial Intelligence as a Proposed Solution

Considering the pedagogical burden involved in developing rich, high-impact tasks for the flipped-classroom model, educators are turning to generative Artificial Intelligence (AI) as a promising aid. The advent of sophisticated large language models (LLMs) presents a new kind of “assistant” for teaching staff. For example, AI tools can serve as “a tool to support the design of creative tasks,” thereby helping instructors navigate designer’s block, accelerate the generation of diverse learning activities, and alleviate the time-intensive nature of task development (Pšenáková et al., 2023).

Beyond this specific claim, a growing body of work supports the notion that generative AI can play a meaningful role in instructional design and flipped environments. For instance, explore how LLM-driven flipped-classroom workflows allow teachers to generate peer-questioning tasks and flipped interactions in large classes (Tan, 2023). Other authors, examine how generative AI advances adaptive and personalized learning by enabling the rapid creation of tailored educational materials and task variants (Guettala et al., 2024).

Within this context, it may be argued that generative AI possesses the potential to function not merely as an isolated tool, but as a systematic support mechanism for educators seeking to design problem-based scenarios, complex case studies, role plays, and collaborative projects. In doing so, these technologies may alleviate the pressures placed upon academic staff, facilitate the creation of more dynamic and varied tasks, and shift the focus of instructional design from initial development toward curation, customisation, and facilitation.

## 1.2 The Research Gap and Objectives

While this potential is both clear and theoretically compelling, a significant gap remains within the empirical research. The educational community lacks sufficient objective data to compare the outputs of an AI-assisted design process with traditional, educator-developed tasks. This raises critical questions regarding whether AI-generated content meets the rigorous pedagogical standards required for the flipped classroom. Furthermore, it remains to be determined whether such technology defaults to lower order thinking, as some critics

contend, or if it can genuinely facilitate the sophisticated, creative tasks essential for higher education.

This paper seeks to address this research gap by employing a quantitative comparative content analysis. The study examines a corpus of 20 tasks developed for a university-level statistics course: 10 designed by the course educator (Human Design; HD) and 10 produced by artificial intelligence (AI-Generated; AG). The research intends to address two primary questions:

1. Is there a significant difference in the quality of HD versus AG tasks based on the core pedagogical criteria of Application, Skills, and Collaboration?
2. Is there a significant difference in the cognitive level (as defined by Bloom's Taxonomy) of tasks designed by humans versus AI?

## 2 Literature Review

The Flipped Classroom model is an implementation of active learning approach (Bonwell & Eison, 1991). The foundational work by Bergmann and Sams (2012) in was quickly adopted by higher education, as it directly addresses the need to move beyond passive information transmission. The flipped classroom model is considered effective because it closely aligns with constructivist learning theories, which emphasize that learners actively construct their own understanding through experience and reflection rather than passively receiving information. This alignment has been repeatedly noted in the literature as one of the theoretical foundations of the flipped approach (Abeysekera & Dawson, 2015; Bishop & Verleger, 2013).

The in-class phase is where this learning is constructed. Research consistently shows that for an FC model to be effective, these in-class activities must be structured, challenging, and collaborative (Strayer, 2012). If students perceive in-class activities merely as conventional homework, motivation is likely to diminish, potentially leading to the failure of the model. Consequently, the onus on the educator to serve as a sophisticated designer of learning experiences is substantial.

### 2.1 Generative AI in Educational Design

The emergence of powerful generative AI, particularly LLMs like GPT-5, has created a new frontier in education. While much public discourse has focused on student use (and misuse), a significant and growing body of research examines AI's role as a tool for educators (Celik et al., 2022; Luo et al., 2025).

AI is being explored as a tool to:

- Generate formative assessment questions.
- Create diverse case studies and "real-world" problems.
- Differentiate instruction by creating variations of a task for different levels.

- Act as a “thought partner” or “sparring partner” for educators to brainstorm new teaching approaches.

Our previous research fit directly into this context, focusing on a high-leverage application using AI to solve the “task design bottleneck” of the Flipped Classroom (Pšenák et al., 2025). The intersection of these two fields, FC pedagogy and generative AI, is new and largely theoretical. Most current literature consists of “how-to” guides, conceptual papers, and anecdotal reports. While optimistic, this body of work lacks the systematic, comparative analysis needed to validate its claims.

It is unknown if AI, when prompted, produces tasks of comparable quality to those designed by an experienced educator. Specifically, there is a concern that AI might excel at “superficial” content generation (Kwan et al., 2025) but struggle to create the complex, “Higher-Order Thinking Skills” (HOTS) tasks involving analysis, evaluation, and creation, which are the hallmark of an effective flipped classroom (López-Villanueva et al., 2024). This study directly addresses this empirical gap.

### 3 Methodology

This study employed a quantitative comparative content analysis of a task corpus (N=20). This methodology was selected as it provides a systematic and objective framework for evaluating the tasks against a predefined set of evaluative criteria (rubrics). The corpus was divided equally into two distinct groups to facilitate a between-groups comparison.

#### 3.1 Corpus Development

A sample of 10 existing in-class activities was selected from the official repository of a university-level statistics course. These tasks were developed by the course instructor (an experienced educator) over the preceding five academic years and constitute a representative cross-section of the activities typically employed within the curriculum.

A corresponding set of 10 tasks was generated using a large language model (specifically, GPT-4). This process was carefully structured to ensure a fair comparison. For each of the 10 human tasks, we first identified its core learning objective (e.g., “Students will be able to perform and interpret a two-sample t-test”).

This objective was then used in a structured prompt: “Act as an expert university-level statistics educator. Design an in-class, collaborative, problem-based scenario for a 50-minute flipped classroom session. The single learning objective for this task is: [Learning Objective]. The task should require students to work in small groups to produce a solution or interpretation.”

This prompting strategy was designed to explicitly request the “creative tasks” that we advocate for.

### 3.2 Instrumentation and Coding

Each of the 20 tasks was independently scored by the same analyst using two instruments.

1. Pedagogical Rubric: A 3-point scale (0 = Absent, 1 = Partially Present, 2 = Clearly Present) was used to score three criteria:

1. Application: Measures if the task requires students to apply knowledge to a new scenario.

Score 2 (Clearly Present): Task requires application in a novel or real-world-like context.

Score 1 (Partially Present): Task is a straightforward procedural application.

Score 0 (Absent): Task is simple recall.

2. Skills: Measures if the task develops a specific technical or soft skill.

Score 2 (Clearly Present): Task explicitly asks for a skill demonstration.

Score 1 (Partially Present): Skill is implied but not required for the final product.

Score 0 (Absent): No skill development is included.

3. Collaboration: Measures if the task necessitates teamwork.

Score 2 (Clearly Present): Task requires a single, integrated group product or solution.

Score 1 (Partially Present): Task suggests discussion but is still completable alone.

Score 0 (Absent): Task is clearly designed for individual work.

2. Bloom's Taxonomy: Each task was classified according to its highest-order cognitive skill, based on the revised Bloom's taxonomy (Anderson & Krathwohl, 2001).

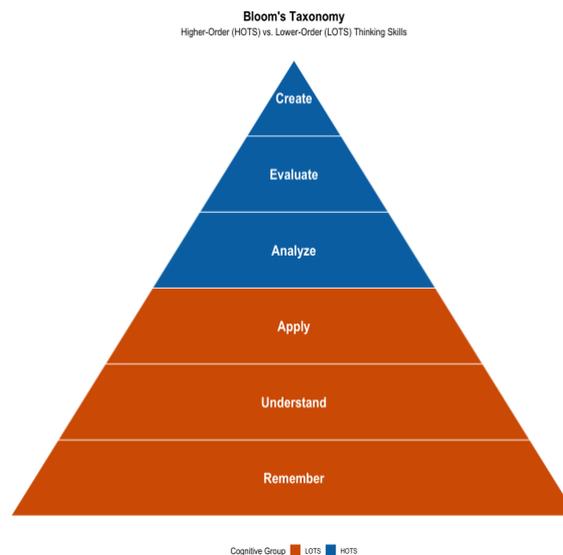


Figure 1: Revised Bloom's Taxonomy (Anderson & Krathwohl, 2001).

For robust statistical analysis, these 6 levels were collapsed into two broader categories: Lower-Order Thinking Skills (LOTS) (Understand, Apply) and Higher-Order Thinking Skills (HOTS) (Analyse, Evaluate, Create). We excluded “Remember” as no tasks fell into this category.

### 3.3 Statistical Analysis

All data were analysed using R, which provides a flexible environment for statistical computing and reproducible research.

**Rubric Data:** The pedagogical rubric scores collected in this study are ordinal in nature and were obtained from a relatively small sample of participants. Given these characteristics, the assumptions for parametric tests (e.g., t-tests) were not met. Consequently, the non-parametric Mann-Whitney U test was employed to compare the distributions of scores for the dimensions of Application, Skills, and Collaboration between the HD and AG groups. This test is well-suited for ordinal data and small sample sizes, as it does not require the assumption of normality and compares the ranks of the scores rather than their raw values (de Winter & Dodou, 2010).

**Bloom’s Taxonomy Data:** To examine the distribution of cognitive levels across groups, a 2x2 contingency table was constructed with the factors Group (HD vs. AG) and Cognitive Level (LOTS vs. HOTS). Pearson’s Chi-squared test was then performed to determine whether the observed frequencies differed significantly from expected frequencies under the null hypothesis of independence. This analysis allows us to assess whether the groups differed in terms of the proportion of tasks targeting Lower-Order Thinking Skills (LOTS) versus Higher-Order Thinking Skills (HOTS). This approach ensures that both ordinal rubric data and categorical cognitive level data are analysed appropriately, providing a robust statistical foundation for comparing instructional outcomes between the two groups.

## 4 Results

The initial analysis of rubric scores showed that both groups produced high-quality tasks that were strong in applying knowledge and developing skills. The mean scores for Application (HD=1.9, AG=1.8) and Skills (HD=1.3, AG=1.5) were very similar.

Group	Mean Application	Mean Skills	Mean Collaboration	Count
HD	1.9	1.3	0.3	10
AG	1.8	1.5	1.8	10

Table 1: Summary of mean rubric scores by group.

The Mann-Whitney U tests confirmed this visual observation. There was no statistically significant difference in Application scores between the HD and AG groups ( $W = 55, p = .58$ ). \*There was no statistically significant difference in Skills scores between the HD and AG groups ( $W = 41, p = .48$ ).

Both the human educator and the AI were equally proficient at creating tasks that required students to apply statistical knowledge and practice relevant skills. In sharp contrast to Application and Skills, the Collaboration metric showed a stark and significant difference. The mean score for the HD group was 0.3, indicating that the human-designed tasks were intended for solo work. The mean score for the AG group was 1.8, indicating that the AI-generated tasks were almost all explicitly designed as group activities. A Mann-Whitney U test found this difference to be highly statistically significant ( $W=6.5, p=.0004$ ).

This result directly supports our prompting strategy. The AI, when asked to create collaborative tasks, was exceptionally effective at doing so, far surpassing the baseline tendency of the human educator.

#### 4.1 Cognitive Level Analysis (Bloom’s Taxonomy)

A primary concern was that AI might “dumb down” content. We analysed the cognitive level of all 20 tasks. The initial 6-level distribution was as follows:

Group	Remember	Understand	Apply	Analyse	Evaluate	Create
HD	0	1	3	3	2	1
AG	0	1	3	2	3	1

Table 2: Frequency of tasks by Bloom’s Taxonomy level.

A Chi-squared test on this 2x6 table failed because the sample size was too small for the number of categories, with many cells having expected counts less than 5.

To resolve this, the categories were collapsed into Lower-Order Thinking Skills (LOTS) and Higher-Order Thinking Skills (HOTS), as described in the methodology. This produced a robust 2x2 contingency table:

Group	LOTS	HOTS
HD	4	6
AG	4	6

Table 3: Frequency of tasks by cognitive group (HOTS/LOTS).

The distribution was identical for both groups. Both the human educator and the AI produced a 60% majority of Higher-Order Thinking Skills tasks. A Pearson’s Chi-squared test confirmed this visual finding, showing no statistically significant difference ( $\chi^2(1) = 0, p = 1$ ).

This approach ensures that both ordinal rubric data and categorical cognitive level data are analysed appropriately, providing a robust statistical foundation for comparing instructional outcomes between the two groups.

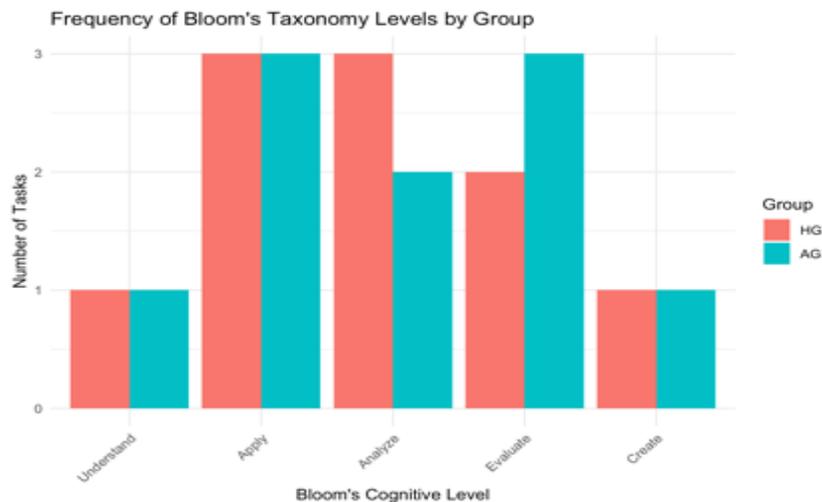


Figure 2: Frequency of Bloom's Taxonomy Levels by Group.

Figure 2 illustrates that the overall distribution of tasks across different Bloom's taxonomy levels is largely similar, with no statistically significant differences observed. This suggests that the Human-Designed and AI-Generated tasks were comparable in their cognitive scope, indicating that AI was able to produce tasks that closely mirror the intended complexity and variety of human-created tasks.

## 5 Discussion

The most valuable finding of the paper is AI's profound and statistically significant strength in generating Collaboration tasks. The human-generated tasks, while pedagogically sound, reflected a common educator tendency of designing solo-work activities (Mean = 0.3). The AI, in contrast, consistently produced tasks (Mean = 1.8) that required group interaction, shared problem-solving, and a single group product.

This directly supports the proposal of this paper that AI can be used to design the "role plays," "problem-based scenarios," and "project activities" that are foundational to the FC model. The AI, unburdened by an educator's established habits or the "worksheet" model of task design, appears to be an exceptional tool for breaking out of this solo-work paradigm.

Our second key finding refutes the common concern that AI may "dumb down" educational content. The AI-generated tasks were indistinguishable from the human-designed tasks in terms of pedagogical quality.

- There was no difference in requiring students to Apply knowledge or develop Skills.

- Crucially, there was no difference in cognitive demand. The AI was just as capable as the human educator at creating tasks that demand Higher-Order Thinking Skills.

This suggests that fears of AI defaulting to lower order, ‘remember’ or ‘understand’ tasks are unfounded, provided the AI is prompted effectively. Our methodology, which explicitly defined the learning objective and context, demonstrates that AI can operate at the higher levels of Bloom’s Taxonomy.

## 5.1 Practical Implications for Educators

The practical implication of these findings is that AI should not be perceived as a replacement for the academic, but rather as a powerful "collaborative assistant". A proposed effective workflow for a contemporary educator employing the Flipped Classroom model is as follows:

1. The educator defines the learning objectives, the specific course context, and any relevant constraints.
2. The educator prompts the AI to generate a variety of creative structures for the in-class activity.
3. The educator then curates, adapts, and performs "human-in-the-loop" editing of the optimal AI-generated output, infusing it with specific domain expertise and an understanding of their students.

This model leverages AI’s while retaining the educator’s essential role as the pedagogical expert. It can save educators significant time in the design phase, allowing them to focus on the in-class facilitation.

## 5.2 Limitations and Future Research

This study has several important limitations:

1. **Sample Size:** The N=20 corpus is small and drawn from a single university course. The findings may not generalize to other disciplines, such as the humanities, or to other educational levels.
2. **Single Coder:** The use of a single coder for the full corpus could introduce bias.
3. **Prompt Dependency:** The quality of AI output is highly dependent on the quality of the prompts. Our structured-prompting method was effective, but different prompting strategies could yield different results.
4. **Existing Tasks:** The HD tasks were pre-existing. A different study design might involve an educator and an AI starting from the same “blank page” to design a task for the same objective.

Future research should expand this methodology to encompass a larger, multi-disciplinary corpus of tasks. However, the most critical progression involves moving from the design phase to empirical implementation. It is imperative that future studies deploy these AI-generated

tasks within a live classroom environment to evaluate their impact on student engagement, the quality of in-class collaboration, and, ultimately, student learning outcomes.

## 6 Conclusion

This study demonstrates that generative AI can effectively support the design of complex, collaborative, and pedagogically sophisticated tasks within the flipped classroom model. The findings indicate that AI-generated tasks demonstrated equivalence to those developed by humans in relation to collaboration, skill development, and higher-order cognitive engagement. Consequently, these results challenge the prevailing concern that the integration of AI may "dilute" or simplify academic content.

The findings suggest a practical workflow wherein academics define core objectives, AI generates a diverse range of task options, and the educator subsequently curates and refines these outputs. This "human-in-the-loop" approach leverages the efficiency and generative capacity of AI whilst retaining the indispensable pedagogical judgement of the academic.

The limitations of this study include the restricted scale of the single-course sample, the reliance on a single-coder analysis, and the inherent dependency on prompt quality. Consequently, it is imperative that future research evaluates AI-generated tasks within authentic classroom environments across a broader range of disciplines. Such investigations are necessary to rigorously assess the impact of these interventions on student engagement, the quality of collaboration, and, ultimately, measurable learning outcomes.

In conclusion, artificial intelligence demonstrates considerable potential as a high-leverage instrument for addressing the "task design bottleneck" inherent in flipped learning. By facilitating the more efficient production of high-quality activities, these tools empower academic staff to overcome resource constraints without compromising pedagogical integrity.

## Acknowledgments

Financial support for this study was provided from the project KEGA 015TTU-4/2024 "The Implementation of the Flipped Classroom Method in Higher Education".

## References

- Abeysekera, L., & Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: Definition, rationale and a call for research. *Higher Education Research & Development*, 34(1), 1–14. <https://doi.org/10.1080/07294360.2014.934336>
- Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman.

- Baig, M. I., & Yadegaridehkordi, E. (2023). Flipped classroom in higher education: A systematic literature review and research challenges. *International Journal of Educational Technology in Higher Education*, 20(1), 61. <https://doi.org/10.1186/s41239-023-00430-5>
- Bergmann, J., & Sams, A. (2012). *Flip Your Classroom: Reach Every Student in Every Class Every Day*. International Society for Technology in Education.
- Bishop, J. L., & Verleger, M. (2013). The flipped classroom: A survey of the research. *ASEE Annual Conference and Exposition, Conference Proceedings*.
- Bonwell, C. C., & Eison, J. A. (1991). *Active Learning: Creating Excitement in the Classroom*. 1991 ASHE-ERIC Higher Education Reports. ERIC Clearinghouse on Higher Education.
- Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The Promises and Challenges of Artificial Intelligence for Teachers: A Systematic Review of Research. *TechTrends*, 66(4), 616–630. <https://doi.org/10.1007/s11528-022-00715-y>
- de Winter, J., & Dodou, D. (2010). Five-Point Likert Items: T Test Versus Mann–Whitney–Wilcoxon. *Practical Assessment, Research and Evaluation*, 15.
- Guettala, M., Bouekkache, S., Kazar, O., & Harous, S. (2024). Generative Artificial Intelligence in Education: Advancing Adaptive and Personalized Learning. *Acta Informatica Pragensia*, 13(3), 460–489. <https://doi.org/10.18267/j.aip.235>
- Kwan, P., Kadel, R., Memon, T. D., & Hashmi, S. S. (2025). Reimagining Flipped Learning via Bloom’s Taxonomy and Student–Teacher–GenAI Interactions. *Education Sciences*, 15(4), 465. <https://doi.org/10.3390/educsci15040465>
- López-Villanueva, D., Santiago, R., & Palau, R. (2024). Flipped Learning and Artificial Intelligence. *Electronics*, 13(17), 3424. <https://doi.org/10.3390/electronics13173424>
- Luo, J., Zheng, C., Yin, J., & Teo, H. H. (2025). Design and assessment of AI-based learning tools in higher education: A systematic review. *International Journal of Educational Technology in Higher Education*, 22(1), 42. <https://doi.org/10.1186/s41239-025-00540-2>
- Pšenák, P., Pšenáková, I., & Szabo, T. (2025). What Happens when we Flip the Class? *R&E-SOURCE*, 283–290. <https://doi.org/10.53349/re-source.2025.is1.a1401>
- Pšenáková, I., & Pšenák, P. (2024). Criteria for Evaluating the Quality of Interactive Didactic Materials. *R&E-SOURCE*, 215–228. <https://doi.org/10.53349/resource.2024.is1.a1257>
- Pšenáková, I., Szőköl, I., & Kováč, O. (2023). The Influence of Innovative Teaching Methods on Student Motivation. *R&E-SOURCE*, 196–203. <https://doi.org/10.53349/resource.2023.is1.a1203>
- Schallert-Vallaster, S., Lavicza, Z., & Vandervieren, E. (2021). Towards Inquiry-Based Flipped Classroom Scenarios: A Design Heuristic and Principles for Lesson Planning. *International Journal of Science and Mathematics Education*, 20. <https://doi.org/10.1007/s10763-021-10167-0>
- Strayer, J. (2012). How learning in an inverted classroom influences cooperation, innovation and task orientation. *Learning Environments Research*, 15, 171–193. <https://doi.org/10.1007/s10984-012-9108-4>
- Tan, C. W. (2023). *Large Language Model-Driven Classroom Flipping: Empowering Student-Centric Peer Questioning with Flipped Interaction* (No. arXiv:2311.14708). <https://doi.org/10.48550/arXiv.2311.14708>