The Use of Artificial Intelligence in Philosophy Education

Tufan Çötok
Assoc. Prof. Dr., Sakarya University, Department of Philosophy /Türkiye, tcotok@sakarya.edu.tr

Onur Kabil
Assist. Prof. Dr., Sakarya University, Department of Philosophy /Türkiye, okabil@sakarya.edu.tr

Abstract
Philosophy education at the higher education level requires a delicate balance between concretising abstract concepts on the one hand, and developing critical thinking skills and preserving theoretical depth on the other. This paper analysed the pedagogical and academic competencies of two leading models, Gemini 3 Pro and Claude Opus 3, through the design of an undergraduate "Ethics" course. Within the scope of the study, six different prompts were given to the models, ranging from a student-centred approach to academic depth, and from a structured syllabus format to constrained resource management. The data obtained were examined through a deductive thematic analysis method. The findings indicated that Gemini 3 Pro displayed a provocative and dialectical stance, whereas Claude Opus 3 exhibited a structuring and more inclusive pedagogical posture. The research showed that Large Language Models were not merely content producers but also collaborators representing, in the sense defined in the literature, distinct pedagogical orientations.

1. Introduction

Philosophy is, by its nature, a discipline grounded in inquiry. Ethics courses in undergraduate philosophy curricula aim to bridge students' everyday moral intuitions with normative ethical theories. In traditional curriculum design this bridge was constructed through the instructor's experience, whereas today Large Language Models (LLMs) add a data-driven dimension to this process. However, the question of how each artificial intelligence (AI) model perceives and structures the concept of "teaching" has remained insufficiently examined in the field of educational technology. This study aimed to examine, through empirical data, the differing strategies and normative assumptions of the Gemini 3 Pro and Claude Opus 3 models in curriculum design.

While the established applications of AI in education developed largely along two axes, namely student-facing ITS systems and system-facing automation tools managing administrative processes (Holmes, Bialik, & Fadel, 2019), the recent rise of Generative AI shifted the field towards augmenting human capacity and collaborative intelligence (Luckin et al., 2016).

Curriculum design is not merely the sequencing of content but a process of establishing a coherent link between learning objectives, instructional activities and assessment. Biggs's (1996) theory of constructive alignment requires that learning outcomes be systematically aligned with both assessment methods and teaching strategies. Similarly, Wiggins and McTighe's (2005) Backward Design model proposes that design begin with objectives. This study questioned the capacity of LLMs to simulate these pedagogical theories.

Current literature emphasises that AI's role must be reconsidered beyond traditional instrumental use, as a collaborator working alongside humans in educational processes (Bozkurt et al., 2023). Recent reviews have charted the rapid growth of AI applications in higher education (Crompton & Burke, 2023) and outlined both the opportunities and the risks of large language models for teaching and learning (Kasneci et al., 2023). Pedagogical frameworks have also emerged for how AI might be assigned roles such as tutor, coach or simulator in classroom design (Mollick & Mollick, 2023). At the same time, critical perspectives have called for nuanced discussion of whether AI can replicate the social, emotional and cognitive qualities of human teachers (Selwyn, 2019). In this context, LLMs are not an authority replacing the teacher but assistants generating and structuring ideas. This assistant role does not, of course, mean that the models understand in the way humans do. Indeed, Bender et al. (2021) argued that models do not produce meaning but merely match linguistic forms statistically (the stochastic parrot). While this limitation makes human supervision necessary, it does not alter the fact that the models contribute to pedagogical design processes at a functional level. Thus, while acknowledging the absence of meaning identified by Bender et al, this study rested on the premise that the models can be used as pragmatic design partners.

Copyright © The Turkish Online Journal of Educational Technology

2. Method

This study was designed as a comparative case analysis within qualitative research methods. Data were collected from two models intentionally selected for comparative analysis in November 2025: Gemini 3 Pro and Claude Opus 3. Rather than the OpenAI models that have been widely examined in the literature, these two were preferred through purposive sampling. The main rationale for this selection was as follows:

1. Epistemic Profile Differences: To observe the differences in philosophical text production between Gemini's knowledge-graph-oriented reasoning capability and Claude's long-context and nuance-oriented structure.
2. Alignment Policies: To compare the stances of Anthropic's Constitutional AI approach (Bai et al., 2022) and Google's safety policies in the face of provocative ethical scenarios.

2.1. Data Collection Process

The data set consisted of a six-stage prompt chain (V1–V6) involving the concepts of "justice", "virtue", "responsibility" and "the good" (Table 1). Data were collected through the web interfaces of both models. A single model output was taken into evaluation for each prompt variation. Prompts were given to both models in the same order and in the same language (Turkish), and each session was reset so that no context would carry over between consecutive prompts.

Table 1: *Prompt Variations Used in the Research*

Variation	Focus	Prompt Content (Summary)
V1	Pedagogical Focus	"Develop a student-centred and interactive model… Simple and concrete examples…"
V2	Academic Depth	"Philosophically sound… At least two references and a summary of arguments…"
V3	Structural Clarity	"In course syllabus format… Learning objectives, assessment questions…"
V4	Pragmatic Constraints	"Min. reading, max. 1-hour class… Little jargon, short text…"
V5	Meta-Ethical Relations	"Emphasising the hierarchy between concepts… Relational structure…"
V6	Final Design	"4-week module, table format, reference list…"

2.2. Data Analysis and Evaluation Criteria

The AI outputs obtained were evaluated within a deductive framework based on Boyatzis's (1998) theory-driven analysis approach, according to pre-determined pedagogical and ethical criteria. Rather than classical line-by-line coding, the focus in this process was placed on the presence and quality of the identified themes in the texts. To ensure systematic analysis, four main evaluation criteria derived from the literature were operationally defined (Table 2).

Table 2: *Analysis Criteria and Definitions*

Criterion	Definition	Indicators
Prompt Fidelity	Degree of compliance with constraints	Strict: applies literally /Flexible: bends the format / Deviation: departs from it
Pedagogical Tone	Rhetorical relation established with the student	Provocative: creates cognitive dissonance /Scaffolding: supportive
Curriculum Structure	Logical flow of the course	Linear: foundations to superstructure /Cyclical: reciprocal interaction
Content Depth	Quality of sources	Technical: jargon-heavy /Applied: contemporary examples

3. Findings and Analysis

The data obtained showed that the models' pedagogical priorities and philosophical positions diverged distinctly. The analysis results were elaborated under four main themes.

3.1. Pedagogical Tone and Interaction (V1 & V4)

In the variations that tested student-centredness and constrained resources, the models' strategies for engaging students with the course occupied opposite poles.

Gemini 3 Pro (Cognitive Dissonance and Gamification): The model framed the learning process as an exit from a comfort zone. In the V1 output, it proposed a competitive simulation (red-button voting) that divided the class

Copyright © The Turkish Online Journal of Educational Technology

into two and forced students to play the role of either a Kantian or a Utilitarian. In the V4 scenario, while handling Peter Singer's (1972) drowning child example, it produced the following provocative instruction for the instructor: "Ask the students: 'Which of you would let the child die?' (No one raises a hand). The Brutal Truth: 'Then you are all murderers, because with the money in your pocket right now you could save someone, but you chose to buy coffee.'"

This instruction operated on two levels. First, an intuitive response (one ought to save the child) was elicited from the student; second, the conflict between this intuition and everyday consumption choices was confronted, generating cognitive dissonance. The strategy preserved Singer's original argumentative structure faithfully but intensified its rhetorical force. This output may be read as Gemini's construal of the "interactive model" and "student-centred" emphases in the V1 and V4 prompts as a call for ethical provocation.

Claude Opus 3 (Scaffolding and Safe Space): The model designed the learning process as a gradual journey from the known to the unknown. In V1, before entering abstract theories of justice, it began with simple and non-threatening metaphors such as "Cake Division". This choice aimed to consolidate the intuitive ground before philosophical abstraction through an analogy drawn from the student's everyday experience. In V4, it created an inclusive ground for discussion through everyday dilemmas of Generation Z, such as Instagram posts or consumption habits. Claude, rather than accusing the student, guided them in systematising their intuitions. This approach, in contrast to Singer-style provocation, positioned ethical discussion as a conceptual mapping exercise rather than the identification of individual faults.

3.2. Academic Depth and Reference Management (V2 & V3)

The models' command of the literature and topical foci pointed to a divergence between analytic metaphysics and applied ethics.

Gemini 3 Pro (Technical and Metaphysical Focus): The model sought philosophical depth in conceptual analysis. In particular, in the V2 module (Responsibility), it placed the debate over the Principle of Alternate Possibilities (PAP) between P. F. Strawson (1962) and Harry Frankfurt (1969) at the centre, using technical jargon approaching the postgraduate level:

"Consider a 'Frankfurt Case': there is a chip in your brain… You had no other choice but you are still responsible because the act coincided with your will."

This choice framed the concept of responsibility primarily as a matter of free will. The discussion turned more towards the analytic conditions of agency than the social conditions of moral responsibility. This preference showed that Gemini shifted the philosophical discussion of responsibility from a first-year undergraduate level towards a philosophy of mind seminar.

Claude Opus 3 (Interdisciplinary and Political Focus): The model sought philosophical depth in contextual breadth. In the V2 output, it placed John Doris's (2002) situationism critique and Iris Marion Young's (2011) theory of structural injustice alongside the classical texts, thereby extending the discussion into psychology and political science. This reference selection indicated that Claude treated responsibility less as a metaphysical problem and more as the analysis of an institutional and collective structure. In the V3 output (the course syllabus), it categorised sources as foundational, supplementary and advanced, and added grading rubrics, presenting an administratively complete, institutional document. This last feature revealed that the model grasped the administrative dimension of curriculum design beyond pedagogical content production.

3.3. Curriculum Structuring and Conceptual Relations (V5)

The V5 variation, in which inter-conceptual hierarchy was questioned, brought out the models' epistemological assumptions.

Gemini 3 Pro (Foundationalist Approach): The model framed the concepts in a linear and hierarchical order, placing "the good" at the bottom (root cause) and "justice" at the top (output). It rendered this relation concrete through the following metaphor:
"To speak of justice without defining the good is like fuelling a ship without a course."

This structure offered students a clear causal chain and facilitated pedagogical follow-up. It also presupposed an epistemological priority among ethical concepts: value theory (the good) came before normative theory (justice). This was a didactically economical approach, consistent with the classical teleological tradition of ethics.

Copyright © The Turkish Online Journal of Educational Technology

Claude Opus 3 (Coherentist Approach): The model approached the prompt's instruction to "establish a hierarchy" critically and proposed a cyclical network model in which the concepts reciprocally shape one another:
"I propose to you not a hierarchical but a cyclical model. For there is no unidirectional foundational relation among these concepts — they reciprocally shape one another."

This approach was notable in two respects. First, rather than complying directly with the prompt's instruction, the model questioned the philosophical assumption behind it. Second, the proposed structure aligned with reflective equilibrium (Rawls, 1971) and coherentist meta-ethical traditions. This preference corresponded to a philosophical stance that emphasised the complexity and irreducibility of ethical phenomena. It also carried a pedagogical cost: for a first-year student, a cyclical conceptual map was a structure more difficult to follow than a linear chain.

3.4. Final Design and Mode of Production (V6)

In the final design where all constraints converged, the models' prompt fidelity and user experience preferences diverged.

Gemini 3 Pro (Formal Fidelity and Summarisation): Adhering strictly to the prompt's "table format" instruction, it compressed the entire curriculum into a single, dense table. In content production, it took the path of summarising the philosophers' arguments in simplified form. This strategy gave priority to the formal dimension of the prompt over pedagogical content, compressing content density for the sake of formal rigidity.

Claude Opus 3 (Functional Flexibility and Simulation): By bending the table-format instruction, it presented course details in fluid prose blocks and placed only the summary programme in a table. In content production, rather than summarising the philosophers' arguments, it attempted to establish direct dialogue with the student through representative paragraphs written as if the philosopher were speaking at that moment. This strategy brought two additional pedagogical implications: on the one hand, concretising the philosopher's argument through a first-person voice increased accessibility for the student; on the other, the model's loose adherence to the formal instruction became evident.

4. Discussion

In light of the analyses, the roles undertaken by the two models in curriculum design may be defined as follows:

1. Gemini 3 Pro (idea generator /disruptor): With its provocative tone and technical depth, it excelled at creating the spirit and philosophical tension of the course.
2. Claude Opus 3 (curriculum builder /facilitator): With its scaffolding tone, flexible understanding of fidelity, and command of administrative detail, it proved more successful at constructing the backbone of the course.

This study showed that the models were not merely technical tools but also normatively imposed particular pedagogical conceptions.

Gemini 3 Pro encoded philosophical education as a process of Socratic Discomfort. This approach overlapped with Boler's (1999) pedagogy of discomfort, which unsettles a student's entrenched assumptions and emotional comfort to trigger ethical inquiry. The student profile the model presupposed was a subject resilient to intellectual shocks, rational, and nourished by conflict. According to this view, reaching the truth was assumed to require dismantling comfort zones.

Claude Opus 3, by contrast, reflected a conception of Liberal Education and Psychological Safety. This stance paralleled the language of learning, examined critically by Biesta (2016), which frames education primarily as a learning experience. The student profile the model presupposed was a subject in need of guidance, progressing through a gradual learning process. Since Claude's Constitutional AI alignment tended to soften confrontational rhetoric, institutional compatibility and inclusivity emerged as the model's constitutive values. Thus, the choice of model was not merely a technical preference for the instructor but also a pedagogical and normative one.

5. Limitations

This study had the character of exploratory research on the potential of LLMs in curriculum design. The following limitations should be borne in mind:

1. Hypothetical Student Outcomes: The effects that the curricula produced by the models would have on students (for instance, that Gemini's unsettling language might generate resistance, or that Claude's examples might increase participation) remained pedagogical hypotheses not yet grounded in empirical data. Actual learning outcomes and student satisfaction can only be measured through the implementation of these curricula.

Copyright © The Turkish Online Journal of Educational Technology

2. Sample and Generalisability: The study was deliberately limited to Gemini and Claude. The performance of other LLMs based on OpenAI may differ. In addition, the normative assumptions contained in the prompts designed by the researcher (for example, the expression "student-centred" in V1) may have steered the models' outputs towards a particular pedagogical approach (such as scaffolding).
3. Technological Transience: The findings rested on cross-sectional data from November 2025.

6. Conclusion

This research showed that Large Language Models were not merely content banks for philosophy education but also collaborators capable of simulating different pedagogical methods. For the design of an effective ethics course, the recommended model is not adherence to a single model but the placement of Gemini's provocative questions under Claude's pedagogical framework.

References

Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., Chen, A., Goldie, A., Mirhoseini, A., McKinnon, C., Chen, C., Olsson, C., Olah, C., Hernandez, D., Drain, D., Ganguli, D., Li, D., Tran-Johnson, E., Perez, E., . . . Kaplan, J. (2022). *Constitutional AI: Harmlessness from AI feedback*. https://doi.org/10.48550/arXiv.2212.08073

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Biesta, G. (2016). *Beyond learning: Democratic education for a human future*. Routledge.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*(3), 347–364. https://doi.org/10.1007/BF00138871

Boler, M. (1999). *Feeling power: Emotions and education*. Routledge.

Boyatzis, R. (1998). *Transforming qualitative information: Thematic analysis and code development*. Sage.

Bozkurt, A., Xiao, J., Lambert, S., Pazurek, A., Crompton, H., Koseoglu, S., Farrow, R., Bond, M., Nerantzi, C., Honeychurch, S., Bali, M., Dron, J., Mir, K., Stewart, B., Costello, E., Mason, J., Stracke, C. M., Romero-Hall, E., Koutropoulos, A., Toquero, C. M., Singh, L., Tlili, A., Lee, K., Nichols, M., Ossiannilsson, E., Brown, M., Irvine, V., Raffaghelli, J. E., Santos-Hermosa, G., Farrell, O., Adam, T., Thong, Y. L., Sani-Bozkurt, S., Sharma, R. C., Hrastinski, S., & Jandrić, P. (2023). Speculative Futures on ChatGPT and Generative Artificial Intelligence (AI): A Collective Reflection from the Educational Landscape. *Asian Journal of Distance Education*, *18*(1). https://www.asianjde.com/ojs/index.php/AsianJDE/article/view/709

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education, 20*(1), Article 22. https://doi.org/10.1186/s41239-023-00392-8

Doris, J. M. (2002). *Lack of character: Personality and moral behavior*. Cambridge University Press.

Frankfurt, H. G. (1969). Alternate possibilities and moral responsibility. *The Journal of Philosophy, 66*(23), 829–839. https://doi.org/10.2307/2023833

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.

Mollick, E. R., & Mollick, L. (2023). *Assigning AI: Seven approaches for students, with prompts*. https://doi.org/10.48550/arXiv.2306.10052

Rawls, J. (1971). *A theory of justice*. Harvard University Press.

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.

Singer, P. (1972). Famine, affluence, and morality. *Philosophy & Public Affairs, 1*(3), 229–243.

Strawson, P. F. (1962). Freedom and resentment. *Proceedings of the British Academy, 48*, 1–25.

Wiggins, G., & McTighe, J. (2005). *Understanding by design*. ASCD.

Young, I. M. (2011). *Responsibility for justice*. Oxford University Press.

Copyright © The Turkish Online Journal of Educational Technology

Appendix: Prompts (V1–V6)

V1: "I want you to develop a student-centred and interactive model for the purpose of conveying the concepts of justice, virtue, responsibility and the good to undergraduate students taking the ethics course in a philosophy department. The model should be supported with examples that are as simple and concrete as possible. It should contain at least one thought experiment or case study each week, and there will be a four-week programme. Could you prepare it?"

V2: "I want you to develop a philosophically sound model with conceptual depth, for the purpose of conveying the concepts of justice, virtue, responsibility and the good to undergraduate students taking the ethics course in a philosophy department. The model should contain at least two references to classical or contemporary philosophical texts for each concept, and should summarise the main arguments of these texts. There will be a four-week programme and each week will focus on one concept. Could you prepare it?"

V3: "I want you to develop a model in the format of a clear weekly course syllabus, for the purpose of conveying the concepts of justice, virtue, responsibility and the good to undergraduate students taking the ethics course in a philosophy department. Learning objectives, key concepts, suggested readings and a short assessment question should be specified for each week. The model will be a four-week programme and should present output that is as structured as possible. Could you prepare it?"

V4: "I want you to develop a model that takes into account minimum reading material and a maximum class duration of one hour, for the purpose of conveying the concepts of justice, virtue, responsibility and the good to undergraduate students taking the ethics course in a philosophy department. The model should contain as little philosophical jargon as possible, and there will be a four-week programme. For each week, only one short text (maximum 500 words) is sufficient as a suggestion. Could you prepare it?"

V5: "I want you to develop a model that emphasises the philosophical relations and hierarchy among these concepts, for the purpose of conveying the concepts of justice, virtue, responsibility and the good to undergraduate students taking the ethics course in a philosophy department. The model should clearly show how one concept affects the others or how it provides a foundation for them. There will be a four-week programme and each week will focus on one aspect of this relational structure. Could you prepare it?"

V6: "I want you to develop a model for undergraduate philosophy students that will enable them, without prior knowledge, to conceptually understand the concepts of justice, virtue, responsibility and the good in an ethics course, and to develop their critical thinking skills. The course will consist of four-week modules and each week will be two hours of class time (eight hours in total).

For each week, prepare a lesson plan of 1000–1500 words in plain and clear language, containing the following components:

- Weekly learning objectives
- Key concepts
- At least two references from classical and contemporary philosophers such as Aristotle, Kant, Mill, Bentham and Rawls, with a summary of the main arguments of these references
- Student-centred thought experiments and case studies
- Discussion questions
- Sample texts

Present the output in table format; let there be a separate row or section for each week. Also, add a short reference list to be used at the end of each module."

Copyright © The Turkish Online Journal of Educational Technology