AI in Education: Risks and Quality Assurance of Pedagogical Practice

Kaja Fjørtoft Ystgaard[1][0000−0001−8427−5033]

Norwegian University of Science and Technology, Trondheim, Norway
kaja.ystgaard@ntnu.no

Abstract. Generative artificial intelligence (GenAI) confronts higher education with a fundamental pedagogical challenge. Current pedagogical practices fail to address how students can use GenAI while keeping their capability to know, reflect and judge, that is, protecting their epistemic autonomy. To mitigate this pedagogical obstacle, we conducted a foresight study, involving both students and educators, to assess which didactic conditions must be in place for students to develop that capability when GenAI is part of learning. Students proved highly AI-literate but traced uncritical use to structural conditions (time pressure, unclear expectations and assessment design) rather than to a lack of knowledge. Educators mirrored this view. They reported insufficient time and a changing role that demands more reflective guidance, more collaboration, redesigned courses and assessments that sustain motivation for preserving students' own sense of knowledge and judgment. To guide this redesign, we offer seven recommendations for pedagogical practice. We argue that quality assurance under GenAI is a matter of institutional condition-setting rather than policing use.

Keywords: Generative AI · Higher education · Epistemic vigilance · Capability approach · Didactics · Anticipatory foresight

1 Introduction

Generative AI has caused a disruption in higher education, but we have largely misdiagnosed it[12]. Current pedagogical interventions tend to focus on what students do with generative AI: whether they use it to cheat, whether they can evaluate its output, whether institutions can detect it, and how the technology affects critical thinking and originality[18]. This transformation of AI use in education has brought new challenges to pedagogical practices and assessment design, where the threats lead to an erosion of disciplinary expertise, a flattening of thinking, and the short-circuiting of human decision-making[9, 6].

To exemplify, when AI is integrated into the practice of learning, teaching, and assessing, we enter a cognitive environment that increasingly supplies ready-made opinions and judgments faster than individuals and groups can form their

© 2026 Copyright held by the owner/author. This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

own[9, 11]. While GenAI is leveraged to enhance accessibility and productivity, it simultaneously poses risks to epistemic autonomy[18], or the ownership of knowledge, while at the same time removing the conditions needed to convert that knowledge into real skills and future capabilities[3].

The pedagogical problem here is not the use of AI in education, but preserving epistemic autonomy while making use of it for learning[18, 3, ?]. The intention of the instruction reflects what we aim to achieve through teaching and pedagogical methods, and the rationales behind them. *Epistemic autonomy* refers to the students' ability to develop a sense of what they are looking for and what a good answer looks like, based on their own, the educators', and the institution's terms[10]. To illustrate in the context of AI and learning, *epistemic vigilance* refers to students' ability to assess whether the AI's knowledge claims are credible, valid, and relevant to the context[20]. This is increasingly regarded as a core component of advanced AI literacy within education and research.

Prior work has not combined an epistemic risk diagnostic with an anticipatory design session[16]. An epistemic risk diagnostics refers to assessing students' epistemic autonomy and vigilance, and what didactic conditions shape their use of AI for educational purposes.

2 Background and Related Work

2.1 Epistemic Risks of Generative AI in Higher Education

Empirical work is increasingly recognizing emerging epistemic risks when using AI for learning. In a pedagogical context, the most significant epistemic risk from conversational AI lies not in inaccuracy or intentional deception, but in something more fundamental[14]. Issues relate to how AI is integrated into didactic decisions, actions and processes, where the purpose (solidarity), depth (co-determination), and agency (self-determination) are removed[10, 9]. The integration of artificial intelligence fundamentally alters how knowledge is produced, validated, and authorized[9].

As an example, among the most severe repercussions of offloading to AI when learning is the weakening of the cognitive muscles. Those who used LLMs throughout to study had weaker understanding, lower recollection, and reduced ability to apply new concepts on their own than those who used them only at later stages[11]. Other unaddressed risks with the use of AI in pedagogical practices relate to a lack of critical thought and exaggerated confidence in the output[21]; dependence, addiction, and not being able to choose to do the task without the tool[2]; a lack of oversight and of the ability to judge what is missing[4]; the homogenization of our thoughts, i.e. our ideas "revert to the mean" or collapse into the most popular or average interpretations[19]; and outside influence from corporate and foreign actors without accountability[9].

Accordingly, the conditions for translating didactic theory into practice need to factor in epistemological displacement, and introduce interventions that target protection from displacement in didactic planning[10].

2.2 Pedagogical Responses: From Detection to Redesign

To address these issues of epistemic risks such as epistemological displacement, in pedagogical practices and assessments, interventions have focused on either AI as a critical-thinking scaffold (support that develops capability and then fades) [22] or protection from AI as offloading or substitution (support that replaces the cognitive work or the epistemological framing permanently)[11].

Whether AI acts as scaffold or substitute depends on the pedagogical design i.e. the didactic decisions, actions and processes[12], where existing research has focused on interventions such as detection regimes[5], and policing AI use for learning. Student experiences of these regimes are less studied than faculty views[5]. Other strands of research have looked at assessment transformation as a part of the redesign and recommendation[13]. These challenges have brought on new ideas around re-designing towards proactive, pedagogically grounded, and equity-focused governance embedded in pedagogical practice[16].

3 Method

3.1 Foresight and Advanced AI Literacy

This study contributes to the quality assurance of pedagogical practice under AI: identifying the conditions that must be in place for pedagogical practice to deliver what it promises.

The methodology we have applied is rooted in advanced AI literacy[1] and foresight design innovation[8]. The foresight methodology allows for comparison across the perspectives of students, educators, and the researchers[8]. Our approach consists of two research stages used to compare a critical assessment of existing conditions in pedagogical practices (critical awareness, independent sense-making, and understanding of AI use for learning)[17] with future anticipatory didactics (with the main focus on pedagogical methods and strategies[10, 8]).

1. Critical diagnostic evaluation of the student perspective
2. Futures design workshop with educators and experts in pedagogy, cybersecurity, AI, and design

Research Design. We applied a capability approach[17] combined with an epistemic vigilance assessment[6], which helped assess how students' capacity to think, question, and reach their own understanding, and their ability to choose based on their epistemic autonomy, and how they are affected when using AI for studying and learning.

Session 1 (90 minutes) was a focus group discussion with 6 students, all fifth-year engineering master students at a large university in Norway. They were recruited through emails contacting the entire student base, and participants signed up based on interest. Session 2 (90 minutes) was a futures design workshop with 5 educators with multi-disciplinary expertise related to cybersecurity, AI and pedagogical/didactic design.

Both datasets were analyzed together using framework analysis[15], with a shared deductive codebook derived from the study's two theoretical lenses: the capability approach (codes: expand, restrict, convert, flourish; functioning vs. capability) and epistemic vigilance (codes: AI literacy, task clarity, source criticism, vigilant use).

Session 2 followed the phases Sense, Prototype and Embed on a Miro board. Session 2 was analyzed from the board artifacts of the session, each element carrying an identifier (e.g., A-F4, B1, C-C16). The Session 1 transcript was checked against the audio recording, and quotes are our translations from Norwegian. Participants gave informed consent, and names in the raw data were pseudonymized.

4 Results

We report the session 1 student findings, the session 2 educator workshop results, and the cross-session mapping between them.

4.1 Student Perspective

Figure 1 maps the themes of Session 1. There is a clear gap between what students report they know about epistemic risks and what they actually verify in educational practice. For instance, they are highly aware of risks, but use AI for most tasks. Most student participants, reflected on losses in the capacity to think, to reach in-depth understanding and achieve clarity on tasks while using AI for studying. Only in certain situations did they demonstrate clarity on tasks. Especially, when working on larger assignments, such as the master thesis, but not for low-effort, concrete and mandatory written assignments. They traced uncritical use to time pressure ("AI or nothing"), mandatory low-relevance courses, and assessment designs that reward AI use, such as a home exam where a trained model could produce the answer. Their AI literacy was high and operational (i.e., next-word prediction, hallucination, sycophancy). The critical distinction the group itself drew was between AI as tool (functioning: things get done) and AI as something that develops or erodes them (capability).

Several students explicitly locate the expansion in learning not in the AI's output but in the meta-skill they built steering it. This is evidence that epistemic vigilance has been learned and that many students have acquired this skill unsupported. However, in this case epistemic vigilance was only applied conditionally: it weakened when prior knowledge was thin, and students said they had built it through early AI models failing, which they doubted later cohorts would experience. Epistemic vigilance in this data is therefore a conversion outcome, depending on disciplinary grounding and a calibration experience, not a pre-existing trait.

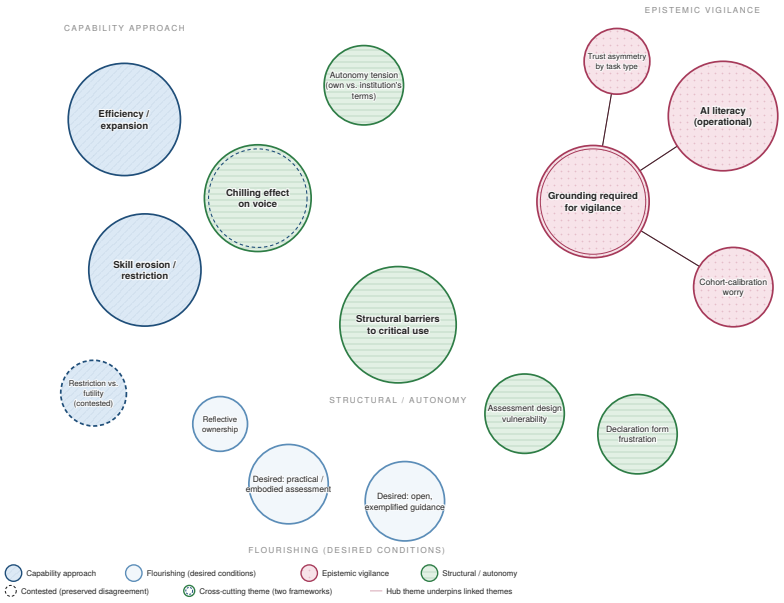


Fig. 1. Thematic salience map of the Session 1 focus group.

4.2 Educator Perspective

In the first workshop phase (sense), educators recognized the student findings as their own: the constraints "also exist on the teacher side". Both educators and students agreed on the need to obtain a solid disciplinary base, as the precondition for epistemic vigilance. During the second phase (prototype) they produced fourteen ideas on discipline-specific literacy, explicit expectations, reflective assessment (AI logs, why/how declarations), motivational course design, and time. In the final stage (embed), pedagogical interventions were grouped between those didactics that educators can easily implement, from those gated by the institution: teacher's capacity and willingness to change role were gated by time, training and slow course development.

5 Recommendations for Pedagogical Practice

The most important recommendation occurred independently in both sessions: both educators and students identified the precondition for epistemic vigilance in disciplinary grounding, course-specific guidance and reflective declaration to sustain students' own sense of knowledge and judgment. To illustrate, anchored in both sessions, the background knowledge required to recognize when AI output fails was a necessary condition for safeguarding epistemic autonomy when AI is part of learning. As such, not being able to know the epistemological disciplinary frames for the subject and the specific context is the most pressing issue

Table 1. Recommendations, graded by agreement across perspectives. A: both educators and students agree; B: partially supported; C: researcher-identified, no educator counterpart.

	For students to...	Practice needs...	First step	Gr.
R1	judge AI output reliably	disciplinary grounding before AI delegation	sequence threshold courses so foundations are built and assessed first	A
R2	use AI critically, on own and the institution's terms	course-specific guidance with examples of good use	write expected AI use into course descriptions	A
R3	reflect on and own their AI use	assessment that trains reflection, not compliance	replace the check-box declaration with a why/how reflection or AI log	A
R4	genuinely choose when to use AI	courses where doing the work oneself is worth choosing	audit mandatory low-effort assignments that create shortcut pressure	A
R5	demonstrate embodied understanding	assessment pitched above what AI automates	pilot one course with an oral or practical component	B
R6	calibrate trust in AI	a designed calibration experience	add an AI-failure exercise to first-semester introductions	B
R7	keep their own academic voice	an enforcement climate that does not penalize authentic style	review detector-adjacent practice; state that style is not evidence of AI use	C

to address[4, 14, 6]. Without factoring in this protection from epistemological displacement, it can be difficult to identify what is missing, what is left behind, or what is not brought forward in the learning process[7, 6].

On the other hand, four student gaps received no educator response: students adapted their writing to avoid AI detection ("chilling effect on voice"), AI agents completing whole assignments, and unequal access to tools. To illustrate, students described changing their own writing (i.e., dropping em-dashes, simplifying their language) to avoid it from looking AI-generated. It can seem as the apparatus meant to protect learning using detection regimes[5], thus restricts the student's own voice. Most students held restriction to be futile, while a minority held that AI is "too strong" and prohibition necessary.

With this in mind, students identified what could work instead: a lecturer starting with a trusted base of texts, open AI use and availability; course-specific guidance; and assessment requiring embodied understanding (oral exams, labs). To redesign didactics a pilot in a single course combining a raised ambition level with an oral or practical component was presented as a partially agreed

and shared solution. Moreover, for students to calibrate trust in AI—which incoming cohorts will not develop accidentally—pedagogical practice needs a designed calibration experience. Both groups named this gap; the intervention is one demonstrative AI-failure exercise on realistic course material.

6 Discussion and Limitations

Read together, nearly every condition the educators formulated ultimately depends on teacher time, training, and willingness to change role—resources the educators reported lacking for the same structural reasons students lack theirs. The conversion barriers of the capability approach[17] are, in this study, symmetric: the conditions that prevent students from converting AI access into genuine epistemic vigilance are the conditions that prevent educators from converting pedagogical insight into didactic redesign. Advice addressed to teaching practice alone would reproduce the problem: some changes are course-level, while time, role change and the enforcement climate need program or institutional decisions[7, 9].

The study is small and qualitative: six self-selected engineering students at one institution, may not represent incoming cohorts. The recommendations were not member-checked and validated in a follow up session. The researcher acted as facilitator, note-taker and analyst, which shaped what was probed and noted.

7 Conclusion

For students to become people who know, reflect and judge, on their own and the institution's terms, while using GenAI, educators and students identify disciplinary grounding before delegation, course-specific guidance, assessment that trains reflection, and courses worth doing with intrinsic motivation for preserving students' own sense of knowledge and judgment. These didactics are gated by institutional practice, so quality assurance of pedagogical practice under AI depends on institutional condition-setting (time, resources and the teacher's changing role) as much as on rules for student use.

References

1. Almatrafi, O., Johri, A., Lee, H.: A systematic review of ai literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). Computers and Education Open **6**, 100173 (2024). https://doi.org/10.1016/j.caeo.2024.100173
2. Basol, M.: Designing ai for humans: preventing manipulation and protecting digital agency (2025). https://doi.org/https://doi.org/10.1093/9780198972877.003.0054
3. Bjork, E.L., Bjork, R.A., et al.: Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. Psychology and the real world: Essays illustrating fundamental contributions to society **2**(59-68), 56–64 (2011)

4. Corbin, T.A., Walton, J.: The missing story of genai summarisers: a critical research agenda. Higher Education Research & Development **44**(7), 1655–1668 (2025)
5. Deep, P.D., Edgington, W.D., Ghosh, N., Rahaman, M.S.: Evaluating the effectiveness and ethical implications of AI detection tools in higher education. Information **16**(10), 905 (2025). https://doi.org/10.3390/info16100905
6. Efimova, E., Nygren, T.: Classroom discussions of social issues in the age of generative ai: Epistemic vigilance against bias and bullshit. The Journal of Social Studies Research **50**(2), 85–97 (2026)
7. Ferdman, A.: Ai deskilling is a structural problem. AI & SOCIETY **41**(4), 3001–3013 (2026)
8. Floridi, L., Strait, A.: Ethical foresight analysis: what it is and why it is needed? Minds and Machines **30**(1), 77–97 (2020)
9. Hartzog, W., Silbey, J.M.: How ai destroys institutions (2025)
10. Klafki, W.: Characteristics of critical-constructive didaktik. Didaktik and/or curriculum: An international dialogue pp. 307–330 (1998)
11. Kosmyna, N., Hauptmann, E., Yuan, Y.T., Situ, J., Liao, X.H., Beresnitzky, A.V., Braunstein, I., Maes, P.: Your brain on chatgpt: Accumulation of cognitive debt when using an ai assistant for essay writing task. arXiv preprint arXiv:2506.08872 **4** (2025)
12. Long, D.Y., Wang, S., Md Rashid, S., Lu, X.T.: Artificial intelligence in higher education: A systematic review of its impact on student engagement and the mediating role of teaching methods. In: Frontiers in Education. vol. 10, p. 1648661. Frontiers (2026)
13. Martin, A.F., Tubaltseva, S., Harrison, A., Rubin, G.J.: Participatory co-design and evaluation of a novel approach to generative AI-integrated coursework assessment in higher education. Behavioral Sciences **15**(6), 808 (2025). https://doi.org/10.3390/bs15060808
14. Maynard, A.D.: The ai cognitive trojan horse: How large language models may bypass human epistemic vigilance. arXiv preprint arXiv:2601.07085 (2026)
15. Ritchie, J., Spencer, L.: Qualitative data analysis for applied policy research. In: Bryman, A., Burgess, R.G. (eds.) Analyzing Qualitative Data, pp. 173–194. Routledge, London (1994)
16. Saito, T.: Redefining agency: A capability-driven research agenda for generative AI in education. Education Sciences **16**(1), 155 (2026). https://doi.org/10.3390/educsci16010155
17. Sen, A.: Inequality reexamined. Oxford university press (1992)
18. Slimi, Z.: A systematic critical review of generative ai's impact on authorship, pedagogy, and integrity (2023–2025). In: Frontiers in Education. vol. 11, p. 1769680. Frontiers Media SA (2026)
19. Sourati, Z., Ziabari, A.S., Dehghani, M.: The homogenizing effect of large language models on human expression and thought. Trends in Cognitive Sciences (2026)
20. Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., Wilson, D.: Epistemic vigilance. Mind & language **25**(4), 359–393 (2010)
21. Tully, S.M., Longoni, C., Appel, G.: Lower artificial intelligence literacy predicts greater ai receptivity. Journal of Marketing **89**(5), 1–20 (2025)
22. Vendrell, M., Johnston, S.K.: Scaffolding critical thinking with generative ai: Design principles for integrating large language models in higher education. Computers and Education: Artificial Intelligence p. 100572 (2026)