EVIDENCE OF A COGNITIVE SHIFT IN AI EDUCATION: HOW STUDENTS ARE RETHINKING HUMAN INTELLIGENCE?

Islem Rekik
BASIRA Lab, Imperial-X (I-X) and Department of Computing
Imperial College London, London, United Kingdom
{i.rekik}@imperial.ac.uk, https://basira-lab.com/

ABSTRACT

Perceptions of intelligence influence how learners evaluate and rely on artificial intelligence (AI) systems. Despite rapid advances in AI capabilities, little is known about how sustained exposure to AI tools affects students' valuation of human (or natural) intelligence (HI) relative to artificial (or machine) intelligence. This Blue Sky paper reports a longitudinal classroom response to a poll comparing the perceived importance of AI versus HI at the opening lecture of AI-focused courses between 2020 and 2026, spanning undergraduate and MSc programs in computer science. Responses from 471 students across technical (Machine Learning, Deep Graph Learning) and design-oriented (Design Thinking for AI) courses were analyzed. We identify four recurring phases: **(1) Hype, (2) Distrust, (3) Trust, and (4) Dependency**. Early measurements in 2020 slightly favored AI over HI. From 2024 onward, preferences consistently shifted toward human intelligence across all MSc cohorts, reaching 65% (a 12 percentage-point increase from 2025) in a technical AI course and 90% (a 36 percentage-point increase from 2025) in a design-oriented AI course by 2026. These observations suggest a gradual reappraisal of human intelligence as AI becomes a routine tool that may affect learners' autonomy and epistemic agency. We conclude this perspective paper by offering introspective insights into a cognitive shift from favoring artificial intelligence toward prioritizing natural intelligence.

1 INTRODUCTION

Artificial (or machine) intelligence (AI) has achieved unprecedented performance across perception, language, and reasoning learning tasks and benchmarks (Xiao et al., 2025; Yue et al., 2025; Chow et al., 2025; Shim et al., 2025). In particular, recent advances in large language models (LLMs) have demonstrated strong capabilities in program synthesis and structured problem solving, enabling models to generate executable code, reason over intermediate steps, and iteratively refine solutions. Empirical studies show that these models can solve a wide range of programming tasks by decomposing problems into modular subcomponents and leveraging learned representations of programming patterns (Chen, 2021; Li et al., 2022; Wei et al., 2022). While these results highlight the effectiveness of implicit, sequence-based reasoning, they have also motivated interest in alternative architectures that aim to endow models with more explicit and persistent cognitive structures. In this context, a new breed of graph neural networks (GNNs) has recently emerged, claiming to equip neural systems with higher-level cognitive capacities such as visual and auditory memory through a dual formalization of graph and reservoir computing (Soussia et al., 2025b;a).

These developments suggest a complementary paradigm for structured reasoning and memory, increasing the appeal of such models in *educational settings*, particularly in computer science, where students use them for coursework, code generation, and conceptual clarification. Empirical studies document that students increasingly integrate large language models such as ChatGPT into their academic workflows for writing, programming, explanation, and problem solving across technical disciplines (Bernabei et al., 2023; Kasneci et al., 2023). As students move from abstract expectations about AI to sustained, everyday interaction with generative models, their lived experience may

reshape how they interpret and value the notion of "intelligence." Therefore, *we hypothesize that increased reliance on AI systems can shift students' learning priorities toward acquiring proficiency in AI tools, potentially at the expense of investing time and effort in developing their human intelligence (HI)*, including metacognitive skills–that is, the ability to monitor and regulate one's own cognitive processes (Fleur et al., 2021).

This paper is motivated by an unexpected empirical observation drawn from longitudinal classroom studies in computer science conducted between 2020 and 2026. After several years in which students displayed mixed impressions about the importance of AI versus HI, a marked shift emerged in 2026: **across two distinct AI-focused courses, students independently reasserted the primacy of human intelligence.**

This Blue Sky paper therefore explores the following prompting question: *What changes in student cognition, self-perception, or lived experience with AI precipitated this reversal?* We emphasize that our analysis is exploratory in nature, offering preliminary introspections and interpretive insights rather than quantitative evidence.

2 Cognitive shifts: A Longitudinal Classroom Signal Interpretation

2.1 Prompt and protocol

Since 2020, at the start of multiple AI-focused courses in computer science (e.g., Machine Learning, Deep Graph Learning, and Design Thinking for AI), MSc and UG (undergraduate) students answered anonymously the following question:

We should invest more effort in improving:
machine intelligence *or* *human intelligence*

Responses were collected in the opening lecture, before technical instruction, to measure *pre-instructional* beliefs. Table 1 summarizes students' responses to this question across AI-focused courses between 2020 and 2026, reporting the relative preference for investing in HI versus AI. **Fig** 1 illustrates the cumulative trend in student preferences for investing in HI versus AI. Each point corresponds to a distinct course offering and is annotated by year and acronym.

In 2020, preferences are mixed, with AI favored in a technically oriented MSc course and HI slightly favored in an undergraduate data course. By 2024, HI is consistently preferred in an advanced technical MSc course (Deep Graph Learning). In 2025, preferences remain near parity but continue to slightly favor HI across both technical and design-oriented AI courses. In 2026, a pronounced shift toward HI emerges, most notably in a Design Thinking for AI course, where 90% of students favor investment in human intelligence.

Although the observations are sparse and not population-representative, they reveal a clear directional shift in expressed preferences over time. We interpret this pattern as suggestive of a broader cognitive and self-perceptual reorientation, in which students gradually re-evaluate the role and value of human intelligence as their direct, sustained experience with AI systems increases.

3 Interpreting the Four Phases of Student Perceptions of Intelligence

Fig 1 reveals a clear directional change in how students value human versus artificial intelligence over time. Because this pattern is derived from a single forced-choice question administered to classroom cohorts, it should not be interpreted as a population-level estimate. Instead, we treat it as an *interpretive signal* that motivates reflection on how sustained interaction with generative AI may influence students' self-perceptions, learning priorities, and epistemic values. The four phases annotated in the figure are therefore not proposed as universal or sequential psychological stages, but as conceptual lenses for organizing these shifts, which may overlap or reoccur as students' experience with AI deepens. We detail the four hypothesized stages below.

Table 1: Student preferences for investing in human intelligence (HI) versus artificial intelligence (AI), collected at the opening lecture of AI-focused courses between 2020 and 2026 in computer science. ITU: Istanbul Technical University. ICL: Imperial College London.

Year	Course	Acronym (institution)	#Students	HI (%)	AI (%)
2020	Machine Learning (MSc)	ML (ITU)	51	41	59
2020	Learning from Data (UG)	LfD (ITU)	89	53	47
2024	Deep Graph Learning (MSc)	DGL (ICL)	106	58	42
2025	Deep Graph Learning (MSc)	DGL (ICL)	113	53	47
2025	Design Thinking for AI (MSc)	DT4AI (ICL)	26	54	46
2026	Deep Graph Learning (MSc)	DGL (ICL)	65	65	35
2026	Design Thinking for AI (MSc)	DT4AI (ICL)	21	90	10

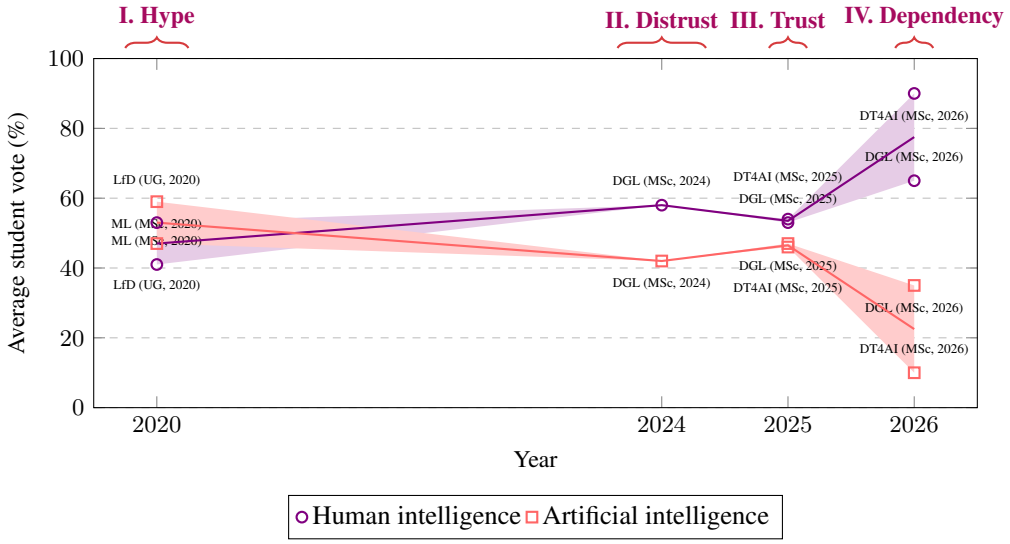

Figure 1: Longitudinal trends in student preferences for investing in human intelligence (HI) versus artificial intelligence (AI) across AI-focused courses from 2020 to 2026. These are based on the student responses to the prompting poll *"We should invest more effort in improving: human intelligence or artificial intelligence."*. From 2024 onward, the yearly mean preference for HI exceeds that of AI, with the largest within-year divergence observed in 2026 due to two course offerings exhibiting sharply contrasting valuations. We interpret this temporal pattern as suggestive of a directional cognitive and self-perceptual shift, in which sustained, hands-on experience with AI systems is accompanied by a re-evaluation of the role and value of human intelligence. Top braces annotate four hypothesized phases (**Hype, Distrust, Trust, Dependency**) to support interpretation of such temporal dynamics.

I–Hype Phase: "The New Hype" ($HI < AI$) In the initial phase, students are highly motivated to explore novel machine learning models and devote substantial intellectual effort to engaging with the emerging hype surrounding artificial intelligence. During this period, AI models are often perceived as intelligent machines, leading students to temporarily deprioritize investment in human intelligence and to downplay its importance (Fig. 1). This shift does not reflect an explicit rejection of human intelligence, but rather a reduced awareness of its central role in the learning process.

II–Distrust Phase: "The Reality Check" ($HI > AI$) As students accumulate direct experience with AI systems, particularly LLMs and generative AI, initial enthusiasm is often tempered by encounters with failure, including hallucinated content, shallow reasoning, and misalignment with evaluative criteria. Such experiences commonly prompt a corrective shift toward skepticism. Verification becomes more salient but also more cognitively demanding, as students must actively monitor and validate AI outputs rather than accept them at face value. In this phase, renewed investment in

human intelligence (HI) is often reactive, driven less by intrinsic valuation of human cognition than by concerns about academic risk, reliability, and accountability.

III–Trust Phase: "The Intellectual Co-Pilot" ($HI \sim AI$) With continued advances in AI tools and their reasoning and problem-solving capabilities, many students develop a more calibrated understanding of AI as an assistive rather than substitutive technology. In this phase, AI is integrated into learning workflows as a cognitive scaffold, supporting brainstorming, clarification, and exploratory reasoning, while responsibility for evaluation and understanding remains firmly with the learner. Crucially, this orientation foregrounds metacognitive regulation: students actively decide when and how to use AI, assess the quality of its outputs, and integrate them into their own reasoning. This phase reflects a pragmatic equilibrium in which artificial and human intelligence are valued for their complementary roles.

VI- Dependency Phase: "The Skill Atrophy Risk" ($HI >> AI$) The final phase reflects a more ambiguous dynamic. As AI systems become deeply embedded in everyday academic practice, some students report difficulty initiating or structuring reasoning without external prompts. Rather than indicating a loss of capability, this pattern suggests a redistribution of cognitive effort, in which generation and organization are partially externalized to the system. Over time, however, such reliance may alter how students perceive their own cognitive competence, leading HI to be valued less for productivity and more for its protective function—maintaining standards of evidence, coherence, and epistemic ownership.

Taken together, these phases provide a lens for interpreting the longitudinal shift observed in **Fig.** 1. The increasing preference for human intelligence can be understood as a cognitive and self-perceptual recalibration shaped by sustained, lived experience with AI systems. The following section offers an introspective discussion of the implications of this shift.

4 Conclusion and Outlook

This Blue Sky paper reports a longitudinal pedagogical signal observed across six years of AI-focused education: a gradual but directional shift in how students value artificial intelligence relative to human intelligence. Early cohorts consistently favored investment in AI, **yet by 2025–2026 this pattern reversed**. In 2026, across two independent cohorts and distinct course contexts, a clear majority of students ranked HI as more important than AI at the beginning of their AI courses. Rather than interpreting this reversal as an abrupt backlash, we argue that it reflects a cumulative cognitive and self-perceptual re-evaluation shaped by sustained, everyday interaction with generative AI systems.

We interpret this trend through four recurring phases of student perception (**Hype, Distrust, Trust, and Dependency**), which together provide a conceptual lens for understanding how exposure to AI reshapes learning priorities over time. Initial enthusiasm frames AI as a shortcut that temporarily obscures the learner's own cognitive role ($HI < AI$). Subsequent encounters with system failures prompt skepticism and a reactive reinvestment in human intelligence ($HI > AI$). With continued use, many students arrive at a calibrated equilibrium in which AI functions as a cognitive scaffold rather than a substitute ($HI \sim AI$). Deeper integration, however, introduces the risk of dependency, where human intelligence is revalued less for efficiency and more for its protective function—judgment, verification, and epistemic ownership ($HI \gg AI$).

The longitudinal shift observed in **Fig.** 1 is consistent with increasing time spent in the latter phases, particularly Trust and early Dependency, suggesting that students may be learning to separate capability from responsibility. This, in turn, suggests a growing recognition that while AI can generate, predict, and optimize, human intelligence remains the locus of accountability and epistemic agency. Another explanation for the post-2023 shift is that students are becoming aware that AI outputs can *shape* human judgement, not merely assist it—an expression of epistemic self-defense[1].

[1]Epistemic self-defense refers to the set of cognitive, metacognitive, and social practices individuals use to protect their beliefs, knowledge, and reasoning processes from deception, error, manipulation, or epistemic harm. It is the ability to defend one's epistemic agency in environments where information is abundant, persuasive, or unreliable.

We emphasize the limits of our evidence. The signal reported here is derived from a single forced-choice question administered to opportunistic classroom cohorts and does not support population-level inference or causal claims. In particular, the prompt *"We should invest more effort in improving: machine intelligence or human intelligence?"* necessarily simplifies a complex, multidimensional relationship into a binary choice, potentially amplifying contrast and obscuring more nuanced or complementary views. Cohort composition, institutional context, and contemporaneous social or technological factors may also contribute to the observed pattern. Nevertheless, the stability of early trends followed by a consistent directional shift motivates systematic investigation. At minimum, these observations suggest that educational settings may serve as early indicators of broader cognitive and cultural changes in *how intelligence is understood and valued*.

Why this matters. If the observed reversal reflects a genuine cognitive reappraisal, it points toward several research directions at the intersection of human-centered AI, education, and human–AI co-learning. These include examining how users' beliefs about AI competence and intent shape reliance and oversight; designing interfaces that scaffold reasoning without displacing epistemic agency; developing hybrid systems that expose structure, rationale, and uncertainty to support human meaning-making; and rethinking curricula to explicitly preserve and assess core human capabilities such as judgment, synthesis, and moral reasoning under pervasive AI assistance.

The observed cognitive shifts are also consistent with broader empirical evidence on when and how human–AI combinations are effective. A large-scale meta-analysis by Vaccaro et al. (2024) shows that hybrid human–AI systems outperform either humans or AI alone *only under specific conditions*– most notably when human judgment, oversight, and contextual reasoning remain actively engaged. **When human agency is diminished or displaced, performance gains from human–AI synergy often disappear or reverse.**

Insight. From this Blue Sky perspective, the goal of education is not to resist AI, but to guide learners toward the Trust Phase while actively preventing a slide into Dependency. Success lies in enabling students to use AI to reach higher levels of abstraction, meta-cognition and creativity without relinquishing the foundational cognitive skills that constitute natural intelligence. Understanding how and when this balance is achieved remains an open and urgent question for future research.

Broader Impact Statement

Understanding how learners' valuations of human and artificial intelligence evolve over time can inform the design of human-aligned AI systems, educational policy, and interaction paradigms that preserve epistemic agency, mitigate overreliance, and support meaningful human oversight in AI-mediated environments. This also challenges prevailing definitions of intelligence and how its various forms are expressed.

LLM Usage Disclosure

In accordance with ICLR policy, we disclose that ChatGPT was used solely for drafting assistance and language refinement. All research questions, interpretations, data collection, and conceptual contributions and formalizations were developed by the authors, who remain fully responsible for the content of this submission.

References

Margherita Bernabei, Silvia Colabianchi, Andrea Falegnami, and Francesco Costantino. Students' use of large language models in engineering education: A case study on technology acceptance, perceptions, efficacy, and detection chances. *Computers and Education: Artificial Intelligence*, 5: 100172, 2023.

Mark Chen. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374*, 2021.

Wei Chow, Jiageng Mao, Boyi Li, Daniel Seita, Vitor Guizilini, and Yue Wang. Physbench: Benchmarking and enhancing vision-language models for physical world understanding. *arXiv preprint arXiv:2501.16411*, 2025.

Damien S Fleur, Bert Bredeweg, and Wouter van den Bos. Metacognition: ideas and insights from neuro-and educational sciences. *npj Science of Learning*, 6(1):13, 2021.

Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, et al. Chatgpt for good? on opportunities and challenges of large language models for education. *Learning and individual differences*, 103:102274, 2023.

Yujia Li, David Choi, Junyoung Chung, Nate Kushman, Julian Schrittwieser, Rémi Leblond, Tom Eccles, James Keeling, Felix Gimeno, Agustin Dal Lago, et al. Competition-level code generation with alphacode. *Science*, 378(6624):1092–1097, 2022.

Jeonghoon Shim, Gyuhyeon Seo, Cheongsu Lim, and Yohan Jo. Tooldial: Multi-turn dialogue generation method for tool-augmented language models. *arXiv preprint arXiv:2503.00564*, 2025.

Mayssa Soussia, Yijun Lin, Mohamed Ali Mahjoub, and Islem Rekik. Coggnn: Cognitive graph neural networks in generative connectomics. In *MICCAI Workshop on Deep Generative Models*, pp. 89–98. Springer, 2025a.

Mayssa Soussia, Mohamed Ali Mahjoub, and Islem Rekik. Multi-sensory cognitive computing for learning population-level brain connectivity. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 521–531. Springer, 2025b.

Michelle Vaccaro, Abdullah Almaatouq, and Thomas Malone. When combinations of humans and ai are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, 8(12):2293–2303, 2024.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.

Wenyi Xiao, Leilei Gan, Weilong Dai, Wanggui He, Ziwei Huang, Haoyuan Li, Fangxun Shu, Zhelun Yu, Peng Zhang, Hao Jiang, et al. Fast-slow thinking for large vision-language model reasoning. *arXiv preprint arXiv:2504.18458*, 2025.

Yang Yue, Zhiqi Chen, Rui Lu, Andrew Zhao, Zhaokai Wang, Shiji Song, and Gao Huang. Does reinforcement learning really incentivize reasoning capacity in llms beyond the base model? *arXiv preprint arXiv:2504.13837*, 2025.