Beyond personalization: autonomy and agency in intelligent systems education

Alie Molly Verghis[1], Binny Jose[2]*, P. R. Sheeja[3],
Sony Mary Varghese[1], S. Mumthas[1] and K. S. Shaiju[4]

[1]Peet Memorial Training College, Mavelikara, India, [2]Department of Health and Wellness, Marian
College Kuttikkanam Autonomous, Kuttikkanam, India, [3]HHMSPB NSS College for Women,
Thiruvananthapuram, Kerala, India, [4]Department of MHTM Marian College Kuttikkanam Autonomous,
Kuttikkanam, India

Introduction

In the era of AI-driven learning, personalization is heralded with much publicity as the harbinger of innovation. Platforms like Duolingo, Coursera, and Khan Academy boast the capability for personalizing content for the learners' needs, pace for maximum learning, and providing intelligent feedback for personal preference. This personalization through algorithm—user behavior, performance metric, and pattern of engagement— has been hailed as the means for learners' empowerment by tailoring learning experiences for learners' unique profiles (Holstein and Doroudi, 2021; Kizilcec and Lee, 2020).

Not all personalization is the same, however. Bellomo (2024) explains the concept in terms of competing visions of pedagogy: extrinsic personalization, which involves aligning external systems with users' behavior, and intrinsic personalization, which focuses on cultivating learners' autonomous capabilities for reflection and meaning-making. Much recent education technology, and much informed by predictive AI, represents the extrinsic model. Though scalable and affordable, these systems might offer a thin conception of personalization—a view with threats to sustain passivity, limit the challenge at the cognitive level, and constrain the student to algorithmically predetermined paths (Cinelli et al., 2020).

This paper engages critically with this extrinsic mode, drawing on Deci and Ryan's (2012) theory of self-determination to argue that over-personalized and gamified systems often undermine autonomy and intrinsic motivation. Students may come to see AI tutors not as guides but as authorities, resulting in emotional outsourcing and reduced epistemic agency (Fakhimi et al., 2023; Placani, 2024).

Rather than expanding learners' intellectual horizons, such systems risk narrowing them—favoring familiarity over exploration. We call for a reclamation of personalization rooted in learner agency, cognitive complexity, and ethical self-direction. This paper advocates for design models that emphasize transparency, override options, and metacognitive engagement.

As educational platforms evolve, it is essential to distinguish between personalization that predicts and prescribes, and personalization that invites curiosity, reflection, and

growth. This paper argues that reclaiming personalization as a learner-driven, ethically grounded, and cognitively expansive process is not only possible—it is necessary.

The constraining nature of personalization

Despite being designed to enhance learning outcomes through customization, algorithmic personalization often narrows rather than broadens the learner's intellectual and developmental horizons. Adaptive learning platforms, while appearing dynamic, frequently operate within rigid algorithmic frameworks that limit rather than expand the scope of engagement. These systems tend to infer a learner's preferences and abilities based solely on historical performance, engagement metrics, or click patterns—criteria that reflect past behavior rather than future potential (Mejeh et al., 2024).

This generates a form of "self-looping" personalization, with the students constantly studying whatever the system identifies as familiar, possible, or high-achieving. These types of feedback loops might give rise to the development of intellectual monocultures, suppress experimentation, critical thinking, and openness to unfamiliar ideas. In short, learners might find themselves trapped in paths selected by the algorithm and reinforce past strengths without ever testing cognitive capacities. As Cinelli et al. (2020) argue in the context of social media, recommendation systems tend to produce echo chambers—an effect mirrored in educational platforms when personalization lacks mechanisms for diversity, challenge, or override.

In addition, few personalization systems incorporate legitimate forms of student agency. Few students are given the opportunity to view how the system creates recommendations or are given the tools with which to challenge or modify them. In this manner, the ideal of empowerment is a technologically mediated form of restriction. This is doubly problematic where systems incorporate gamified features that drive learners toward system-approved behavior through the use of reward mechanics, badges, or progress loops. These frameworks, while powerful at driving engagement, do so at the cost of the loss of intrinsic motivation and autonomy (Deci and Ryan, 2013).

While in ideal situations these systems would promote independent learning with metacognitive reflection, they could actually produce help-dependent learners who rely upon external cueing in order to progress. Fakhimi et al. (2023) and Placani (2024) report that emotionally intelligent AI tutors themselves can generate trust levels that suppress skepticism and critical action. In the absence of openness, these systems are black boxes with learning paths dictated by them without learners' informed consent or understanding.

Lifting these limitations involves shifting away from personalization as prediction and toward personalization as discourse. Systems need to help learners with goal-setting, progress reviewing, and informed decision-making on the paths of learning. Rather than closing down opportunities, personalization must organize exploration and enable consequential learner autonomy.

Subtle nudges: guiding or governing education?

Algorithmic nudges—offering tiny prompts within system design—could steer learners toward desired action, content, or order (Thaler and Sunstein, 2009). In learning spaces, they might show up as badges, gamification rewards, or prompts within the interface that reflect desired progress (Disalvo and Morrison, 2013). While increasing measures of engagement, they could actually divert students away from critical reflection and self-directed learning.

From a psychological standpoint, these nudges could be small extrinsic reinforcers that function more as behaviorist conditioning rather than as teaching for autonomy (Deci and Ryan, 2012). The problem is that we reinforce compliance and speed more than we reinforce curiosity or cognitive struggle.

Not every nudge is the same, however. Platforms like edX and FutureLearn, for instance, allow learners the option of overriding system suggestions, reviewing earlier content and going back, or pacing for the sake of personal aims (Khosravi et al., 2022). These affordances exhibit an alternate design ideal—one by which the nudge is recognized as a dialogical signal and not an imperative. This diversity in design is worth noting lest all adaptive sites be subjected under the same critique.

Moreover, invisibly or without the consciousness of the learner, nudges might undermine epistemic agency—such that learners come to trust the system more than their own judgment. Placani and Fakhimi et al. caution against the danger that emotionally convincing or affectively attuned systems might cause the loss of skepticism and foster over-compliance rather than the use of critical thinking.

Learning design for agency involves incorporating nudges that require reflection and not compliance. The focus should be on creating intentional learning behavior without taking away the learner's freedom to question, vary, or resist system-generated prompts.

Ethical paradoxes of algorithmic personalization

Algorithmic personalization raises important ethical questions—specifically with respect to transparency, consent, and the construction of learner profiles. Behaviorally monitoring systems that change themselves in response potentially compromise learners' abilities for informed decision-making. The problems are more significant where learners are unfamiliar with how their information is used in informing responses for them.

Placani (2024) emphasizes how personalization systems can simulate care, fostering emotional trust in systems that lack actual understanding or accountability. This can create a false sense of rapport—what has been termed emotional outsourcing—where the learner assumes cognitive and emotional guidance from systems designed primarily for behavioral optimization (Fakhimi et al., 2023; Mohanan et al., 2017).

However, ethical issues are not uniformly distributed across all platforms. While many personalization systems function as opaque black boxes, others are actively working to incorporate explainable AI principles, user consent protocols, and transparent design features (Adeniran et al., 2024; Harve et al., 2024; Khosravi et al., 2022). In some instances, learners are even provided access to data visualizations, progress tracking tools, or algorithmic explanations—practices that can foster critical reflection and ownership.

As Cinelli et al. (2020) point out, the opacity of algorithmic systems can also contribute to ideological and cognitive echo chambers. In the educational domain, such structures may shape not just what students learn, but how they learn to think and engage with knowledge.

What is needed is a framework that aligns personalization with educational ethics: fostering clarity, learner empowerment, and shared decision-making. Ethical personalization is possible—but it requires intentionality and accountability at every layer of design and deployment.

Reclaiming autonomy

If personalization is to truly empower learners, it must be rooted in principles that support autonomy, reflection, and self-direction. This shift begins by rethinking how AI systems frame the learner: not as a passive recipient of recommendations, but as an active agent in a dynamic learning process. Drawing on Bellomo's (2024) distinction between extrinsic and intrinsic personalization, we argue that effective personalization must cultivate internal decision-making capacities rather than merely adjusting content to fit past behavior.

One way to reclaim agency is by designing systems that make their operations transparent and navigable. Explainability should not be a hidden feature but a central design principle. When learners understand how and why a system makes certain recommendations, they are better positioned to question, reject, or refine those pathways. Research in explainable AI, such as Adeniran et al. (2024), shows that transparency increases trust without diminishing user autonomy.

No less important is the capability for override. Students need to be able to choose learning paths that suit their aims even if the paths do not suit system projections. Systems that allow learners the option of challenging or tailoring recommendations—in the form of open-ended modules, reflection questions, or learning contracts—allow for space for negotiation between the algorithmic rational and intentionality.

Furthermore, the cultivation of agency is the integration of metacognitive tools into the learning interface itself. Self-test checklists, progress records and dashboards for the learner, or goal-setting prompts can assist in providing a direction and a sense of ownership. As Deci and Ryan (2012) maintain, autonomy is not the elimination of control so much as the introduction of important choice within an enabling context.

Well-designed AI tutors might actually promote learner autonomy rather than undermine it. Placani and Fakhimi give us a caution against AI systems that mask judgment with affective mimicry—empathetically intelligent messages that conceal a lack of epistemic openness. The problem is not how we might make AI tutors more human, but how we might make them more respectful of educational values such as clarity, deliberation, and respect for learner agency.

For this end, we advocate design frameworks grounded in intrinsic personalization. They position growth for the learner before behavioral optimization, foster epistemic modesty for system suggestions, and invite learners into a co-creative discourse with technology. Retaking agency within the intelligent learning space is not an issue of resisting personalization—but of redefining it on human-focused, autonomy-promoting terms.

Beyond classrooms: the ripple effect of personalization

Algorithmic personalization in education reflects a broader societal pattern seen in social media, healthcare, and employment systems. Across domains, personalization filters experience, often limiting exposure to diverse or dissenting perspectives.

On platforms like social media, algorithms create echo chambers that restrict intellectual diversity (Cinelli et al., 2020). In education, similar logic may foster intellectual monocultures, reinforcing students' existing preferences rather than challenging them (Helberger et al., 2018).

There is a counterexample from healthcare: AI systems can personalize therapy successfully but can equally encode systemic biases and constrain explainability (Harve et al., 2024). These are two examples that demonstrate that personalization can diminish autonomy without being accompanied by explainability and responsibility.

Education, thus, plays a distinctive role. It can prototype ethical personalization by integrating transparency and intellectual surprise—injecting the unexpected content or alternative perspectives. Rather than reinforcing preferences, it can instruct students to be critical of algorithmic systems they find everywhere else.

Limitations and future directions

This paper is intentionally conceptual and interpretive. While the arguments draw upon interdisciplinary scholarship in education, AI ethics, and psychology, they are not grounded in empirical fieldwork or platform-specific evaluations. Future research should aim to validate or refine these claims through empirical studies that examine how learners interact with personalization systems in varied contexts.

Mixed-method approaches—including ethnographic observation, platform log analysis, and learner interviews— could provide richer insights into when and how personalization enhances or limits autonomy. Comparative research across platforms would be especially valuable to identify design features that promote intrinsic motivation and learner agency.

Moreover, Bellomo's (2024) framework distinguishing extrinsic from intrinsic personalization offers a valuable theoretical lens for such investigations. Future design-based research could explore

how systems can operationalize intrinsic personalization to support epistemic growth and ethical development in learners.

Conclusion

Personalization in AI-mediated education is not inherently flawed—but its dominant forms often reflect extrinsic, behaviorist assumptions that limit rather than liberate learners. By conflating efficiency with empowerment, many current systems risk narrowing curiosity, discouraging agency, and over-relying on predictive optimization.

This paper has argued for a conceptual reorientation: reclaiming personalization as a learner-driven, ethically grounded, and cognitively expansive process. Drawing on frameworks like Bellomo's intrinsic personalization and Deci and Ryan's theory of self-determination, we advocate for systems that prioritize transparency, override capabilities, and metacognitive reflection.

Ultimately, the goal is not to reject personalization but to redefine it. When AI systems are designed to foster autonomy, embrace complexity, and respect learners as co-authors of their educational journeys, personalization can become a tool for liberation—not limitation.

Author contributions

AV: Writing – original draft, Writing – review & editing, Supervision. BJ: Writing – original draft, Writing – review & editing, Investigation, Conceptualization. PS: Writing – original draft, Writing – review & editing. SV: Writing – review & editing, Writing – original draft. SM: Writing – original draft, Writing –

review & editing. KS: Writing – original draft, Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adeniran, A. A., Peace, A., and William, P. (2024). Explainable AI (XAI) in healthcare: Enhancing trust and transparency in critical decision-making. *World J. Adv. Res. Rev.* 23, 2647–2658. doi: 10.30574/wjarr.2024.23.3.2936

Bellomo, S. T. (2024). Extrinsic and intrinsic personalization in the digital transformation of education. *Ethics Open Dist. Educ.* doi: 10.26034/fr.jehe.2024.6861

Cinelli, M., Morales, G., Galeazzi, A., Quattrociocchi, W., and Starnini, M. (2020). Echo chambers on social media: a comparative analysis. *ArXiv.* doi: 10.48550/arXiv.2004.09603

Deci, E. L., and Ryan, R. M. (2012). Self-determination theory. *Handb. Theor. Soc. Psychol.* 1, 416–436. doi: 10.4135/9781446249215.n21

Deci, E. L., and Ryan, R. M. (2013). *Intrinsic Motivation and Self-Determination in Human Behavior.* Springer Science and Business Media.

Disalvo, B., and Morrison, B. (2013). *A Critique of "Gamification" in Khan Academy.* Available online at: https://consensus.app/papers/a-critique-of-%E2%80%9C-gamification-%E2%80%9D-in-khan-academy-morrison-disalvo/b5bfbb77b04450ea9041e28bd4aa5cbe/ (Accessed July 14, 2025).

Fakhimi, A., Garry, T., and Biggemann, S. (2023). The effects of anthropomorphised virtual conversational assistants on consumer engagement and trust during service encounters. *Austral. Market. J.* 31, 314–324. doi: 10.1177/14413582231181140

Harve, M., Ramamoorthy, S. P., Pati, V., Bainbridge, G., Kankolenski, A., Podder, B., et al. (2024). Reducing racial biases within healthcare applications of

artificial intelligence (AI) with transparency. *J. Sci. Policy Govern.* doi: 10.38126/JSPG 250110

Helberger, N., Karppinen, K., and D'Acunto, L. (2018). Exposure diversity as a design principle for recommender systems. *Inform. Commun. Soc.* 21, 191–207. doi: 10.1080/1369118X.2016.1271900

Holstein, K., and Doroudi, S. (2021). Equity and artificial intelligence in education: will "AIEd" amplify or alleviate inequities in education? *ArXiv.* doi: 10.48550/arXiv.2104.12920

Khosravi, H., Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J. F. L., et al. (2022). Explainable Artificial Intelligence in education. *Comput. Educ. Artif. Intell.* 3:100074. doi: 10.1016/j.caeai.2022.100074

Kizilcec, R., and Lee, H. (2020). Algorithmic fairness in education. *ArXiv.*

Mejeh, M., Sarbach, L., and Hascher, T. (2024). Effects of adaptive feedback through a digital tool–a mixed-methods study on the course of self-regulated learning. *Educ. Inform. Technol.* 29, 1–43. doi: 10.1007/s10639-024-12510-8

Mohanan, R., Stringfellow, C., and Gupta, D. (2017). "An emotionally intelligent tutoring system," in *2017 Computing Conference* (London: IEEE Conference Publication), 1099–1107. doi: 10.1109/SAI.2017.8252228

Placani, A. (2024). Anthropomorphism in AI: hype and fallacy. *AI Ethics* 4, 691–698. doi: 10.1007/s43681-024-00419-4

Thaler, R. H., and Sunstein, C. R. (2009). *Nudge: Improving Decisions About Health, Wealth, and Happiness.* London: Penguin.