

Designing for Persistence: AI Chatbot Architecture, Gender Bias, and Inclusive Design for Computing Higher Education

Alina Berry

Technological University Dublin

alina.berry@mytudublin.ie

Abstract

Recent advances in large language model (LLM) chatbots have led to their increasing use in computing education, particularly for feedback, assessment support, and introductory programming assistance. While these systems offer clear benefits in terms of scalability and personalised learning support, the literature highlights persistent limitations, including shallow pedagogical reasoning and the continued need for human oversight in educational settings. Current approaches increasingly favour retrieval-augmented generation (RAG) architectures and course-specific tutoring systems, which ground responses in curated instructional material and constrain model behaviour to improve reliability and contextual relevance.

Despite these technical advances, there is still limited understanding of how AI-supported learning environments shape student experience, particularly in relation to gender equity, confidence, and persistence in computing. This is significant given persistent underrepresentation of women in computing and growing evidence that AI systems can reproduce and amplify social stereotypes in ways that affect learner engagement and belonging.

This work-in-progress paper synthesises evidence across AI chatbot architecture, gender bias in model outputs and chatbot design, and mitigation approaches, and argues that technical grounding alone does not guarantee equitable learning outcomes. It proposes a socio-technical model for gender-responsive LLM-based AI teaching assistants that uses RAG architecture. The proposed model is centred on six design considerations: *inclusive knowledge architecture*, *stereotype threat mitigation*, *neutral persona design*, *communication functionality*, *participatory co-design*, and *integration within wider intervention ecosystems*.

The paper further argues that AI teaching assistants should be evaluated not only in terms of accuracy and learning gain, but also in terms of their impact on confidence, engagement, and persistence across diverse student groups. It concludes that equity in educational AI is not a post-hoc adjustment, but a property that must be designed into both the knowledge base and the interaction layer of the system.

1. Introduction

The past decade has seen a growing use of conversational AI tools in higher education. In computing disciplines specifically, LLM-powered assistants are being increasingly deployed to answer course questions, give formative feedback, support code review, and provide personalised instruction (Pitts, Hridi, & Narayanan, 2025; Elnaffar, 2025; Bekkar & Chtouki, 2024; Sadat Shanto, Ahmed, & Jony, 2024). The most common promise is scale and personalisation (Baidoo-Anu & Owusu Ansah, 2023; Kasneci et al., 2023; Yigci et al., 2024); the most consistent warning is that fully automated systems miss pedagogical nuance (Sadat Shanto et al., 2024) and need human oversight to avoid shallow learning or integrity problems (Pitts et al., 2025; Jamie, Hajihashemi, & Alipour, 2025).

These concerns are urgent for any learner, but they carry particular weight for a population that research consistently identifies as especially vulnerable to differential AI treatment: women and other underrepresented gender groups in computing (Bender, Gebru, McMillan-Major, & Shmitchell, 2021; Blodgett, Barocas, Daumé III, & Wallach, 2020; Kotek, Dockum, & Sun, 2023). Women account for fewer than 20% of employed computing professionals across Europe (Eurostat, 2024), and retention of women in computing higher education is an equity challenge (Lehman, Newhouse, Sundar, & Sax, 2023). In Ireland, across the 2016/2017 to 2023/2024 academic years, ICT-related programmes had among the

highest non-progression rates for women and non-binary students, at 11% (Higher Education Authority, 2026), compared, for example, with engineering at 8.4%, mathematics/statistics at 8.1% or education at 3.2%. Even chatbots not explicitly designed as gender-specific interventions have been shown to increase confidence among women computing students (Leavitt et al., 2022), and increased confidence in STEM disciplines has been associated with retention (Dennehy & Dasgupta, 2017). AI tools have been shown to encode and reproduce gender bias through their training data, algorithms, and feedback mechanisms (Kotek et al., 2023; Wan et al., 2023; Bender et al., 2021), meaning that tools not deliberately designed with gender in mind risk reinforcing the bias that could affect women's confidence and in turn, their persistence in computing.

While some institutions restricted or banned the use of AI (Yigci et al., 2024; de Fine Licht, 2024), universities cannot realistically prevent students from accessing AI tools through personal smartphones or external networks. Consequently, and considering that, for instance, in the UK 92% of higher education students use AI in their coursework (Freeman, 2025), institutionally supported AI programming assistants may provide a more effective approach by enabling guided, transparent, and pedagogically aligned use of AI for learning, debugging, explaining code, and providing accessible on-demand support.

This paper is structured as follows. Section 2 examines how contemporary educational AI chatbots are constructed. Section 3 reveals the open research gap. Section 4 synthesises the evidence on gender bias in LLM outputs and chatbot design. Section 5 reviews mitigation strategies and proposes six design considerations for equitable deployment in university programming classes, which are then shaped into a suggested socio-technical model. Section 6 recommends evaluation steps for the model. Finally, Section 7 concludes this work and lists next steps.

2. Technical Architecture

For the purpose of this research, AI chatbots, or AI Teaching Assistants (TAs) are defined as text-based conversational agents that interact with students via natural language interfaces (Smutny & Schreiberova, 2026). Two broad types of AI chatbots in education are frequently discussed in the literature: rule-based chatbots and large language model (LLM)-based chatbots (Smutny & Schreiberova, 2026; Kuhail, Alturki, Alramlawi, & Alhejori, 2023; Fetrati, Chan, & Orji, 2025). Rule-based chatbots use predefined guides through fixed dialogue flows and while they were dominant before the widespread adoption of LLM-based approaches, they still remain in active use (Kuhail et al., 2023), including alongside the LLM-based chatbots (Fetrati et al., 2025). Because rule-based systems potentially require substantial manual design to cover the range of support needed in programming education, this paper focuses primarily on LLM-based AI teaching assistants.

2.1. Large Language Models

LLM-based chatbots, which are becoming increasingly popular in educational contexts (Giannakos et al., 2025; Smutny & Schreiberova, 2026), are trained on large amounts of internet-derived data and can be refined using curated instructions with alignment to human expectations through a process known as Reinforcement Learning from Human Feedback (RLHF) (Ouyang et al., 2022). Each stage of that pipeline can introduce bias: the pre-training corpus reflects who produces content on the internet and who is written about (Bender et al., 2021); the fine-tuning data encodes the judgements of the human annotators who labelled it (Casper et al., 2023); and the RLHF process reflects the preferences of a rating workforce whose demographic makeup is rarely reported or disclosed (Casper et al., 2023).

2.2. Retrieval-Augmented Generation (RAG)

The computing education literature increasingly points toward Retrieval-Augmented Generation (RAG) as a defensible architecture for course-specific teaching assistants (Liu et al., 2024; Feng, Liu, & Ghosal, 2024; Wang & Lawrence, 2025; Aliyeva, 2025; Dakshit, 2024). In a RAG system, a student query is matched against a course-specific knowledge base and the retrieved passages are then used to ground the model's response in local instructional material (Liu et al., 2024; Feng et al., 2024; Swacha & Gracel, 2025; Mahajan, 2025). The aim is that the assistant stays tied to the course, the lecture notes, and the instructor's own framing of the material. That grounding is exactly what beginner programming AI tools

were designed to do: approximate a 1:1 teacher-to-student ratio, respond to code snippets and course queries, and support students without replacing the human educator (Liu et al., 2024).

Comparative evaluations point in the same direction. CourseAssist, an LLM-based tutoring system for over 500 computing students (Feng et al., 2024) uses retrieval-augmented generation, user-intent classification, and question decomposition to align responses with course materials and learning objectives, and the authors report that it outperformed a GPT-4 baseline on usefulness, accuracy, and pedagogical appropriateness. Wang and Lawrence’s evaluation of multiple open-source and commercial models shows the same instructor-facing logic: the relevant question is not whether a model can answer, but whether it can answer well enough to justify the maintenance burden of supporting it in a course setting (Wang & Lawrence, 2025). Németh et al. report that a RAG tutor in higher education reduced hallucinations for topics sufficiently covered in the course material, while students used it as help for comparison, study organisation, and active learning rather than as a replacement for the course (Németh, Tátrai, Szabó, & Tamási, n.d.). Aliyeva’s university Computer Science (CS) study (Aliyeva, 2025) reaches a similar conclusion, concluding that RAG-based systems can provide personalised support in programming classes. Dakshit’s faculty-perspectives paper discusses that faculty members acknowledged RAG’s potential as a virtual teaching assistant, but also named the ethical and operational safeguards needed for deployment (Dakshit, 2024).

3. The Open Research Gap

Cao et al. synthesise 98 papers on generative AI teaching assistants and report that RAG-enhanced systems outperform pure generative models on accuracy and learning effectiveness, but remain constrained by answer accuracy, scenario adaptation, and academic-integrity concerns (Cao, Shao, Zhang, & Gao, 2025). Martin et al. make the equity gap even harder to ignore: in their review of adaptive pedagogical systems in CS education, there was little to no data on effectiveness by gender or race, and most papers did not include user demographics (Martin, Ying, Rodríguez, Kahn, & Boyer, 2022). That absence matters because it means we have a growing body of tools with very little evidence about whether they serve all learners equally.

3.1. The gender-equity problem is not separate from the architecture problem

Berry et al. (Berry, McKeever, Murphy, & Delany, 2022) review the literature on women’s retention and recruitment in computing education and classify the literature on gender initiatives that worked into Policy, Pedagogy, Influence and Support, and Promotion and Engagement, while also noting how often intervention studies fail to evaluate their own impact rigorously. Perez-Felkner et al. (Perez-Felkner et al., 2025) extend that logic across 48 studies and show that the effectiveness of interventions depends on intervention domain, strategy, goal, level, and duration. The broader intervention literature supports the same conclusion from multiple angles: mentoring, role models, curriculum innovation, and institutional support repeatedly appear as the most plausible factors for improving women’s retention and participation in computing (Szlavi, Haugland, & Jaccheri, 2024; Thompson & Wu, 2021; Pantic & Clarke-Midura, 2019).

A course-specific RAG-based teaching assistant may inherit the assumptions, examples, pronouns, and omissions of the course material it indexes. The gender gap in computing is partly sustained by a ‘male-default’ culture in which computing is stereotyped as masculine and socially isolating. These stereotypes influence belonging, self-efficacy, participation, and persistence among women in computer science (Beyer, 2014; Yansen & Zuckerfeld, 2014; Osunde, Windall, Bacon, & Mackinnon, 2014). If that RAG-based course corpus reflects a male-default vision of computing, then the assistant will reproduce that vision at scale, quietly and repeatedly, in the very place where students are forming beliefs about whether they belong. The composition of the knowledge base is therefore a design question, not a back-end implementation detail. That point is consistent with the inclusive-computing RAG prototype designed for Spanish-speaking regions, which was built explicitly to support gender-equity curriculum feedback and reported that human reviewers rated its outputs as pedagogically relevant and equity-aware (Ormeño-Arriagada, Herrera, & Queirolo, 2025).

The practical implication is that a RAG-based teaching assistant should not only be judged on answer accuracy. It should also be judged on whether it supports confidence and belonging, reduces friction, and avoids reproducing gendered assumptions already present in the field. That is why the most useful RAG work in this context is the work that couples grounded retrieval with explicit instructional control and with equity-aware evaluation (Liu et al., 2024; Feng et al., 2024; Ormeño-Arriagada et al., 2025; Martin et al., 2022). The architecture is promising, but the equity claim still has to be earned through the prompts, the feedback rules, and the subgroup analysis.

4. Gender Bias

4.1. Bias in LLM Outputs

The evidence that LLMs exhibit systematic gender bias is now substantial. Kotek et al. (Kotek et al., 2023) demonstrated that contemporary LLMs were 3–6 times more likely to assign occupations that stereotypically aligned with the prompted gender, and often justified those choices with internally consistent but factually inaccurate explanations. The study further found that model outputs aligned more closely with social stereotypes than with real labour-market statistics.

Recent work suggests that these patterns persist even in models designed to appear more balanced or inclusive. Chaturvedi and Chaturvedi (Chaturvedi & Chaturvedi, 2025) show that LLMs used in hiring-style tasks favour men in higher-wage occupations and reproduce established occupational stereotypes even when prompted to behave neutrally. Related studies on occupational bias in LLMs similarly find that models continue to associate technical and leadership roles more strongly with men than women (Chen, Raghuram, Mattern, Mihalcea, & Jin, 2025).

4.2. Bias in Chatbot Design

Beyond model outputs, the design of chatbot interfaces and personas is itself a site of gender encoding. Feine et al. (Feine et al., 2020) find that many deployed chatbots, particularly in customer service and support contexts, are assigned feminine personas, reflecting broader cultural assumptions linking helpfulness, emotional labour, and service work with women.

Research on stereotype threat, the anxiety induced when a learner fears confirming a negative group stereotype (Steele & Aronson, 1995), is directly relevant to computing education. Even subtle environmental cues could negatively affect confidence and performance for women in technical domains (Murphy, Steele, & Gross, 2007). An educational chatbot that defaults to masculine assumptions about programmers, or frames difficulty as evidence of lack of aptitude rather than limited prior exposure, may silently reinforce those dynamics across repeated interactions.

5. Mitigation Strategies and Design Considerations

5.1. Bias Mitigation Across the Pipeline

Gallegos et al. (Gallegos et al., 2023) frame bias mitigation as something that has to happen across the entire machine learning pipeline, rather than as a single post-hoc fix. This includes pre-processing steps such as dataset curation, balancing, and counterfactual augmentation; in-training approaches such as adversarial debiasing and fairness-constrained objectives; intra-processing methods like activation steering or attention-based interventions; and post-processing techniques including guardrails, safety classifiers, and constrained decoding.

Related survey work highlights similar practical levers used in real deployments, including curated training data, counterfactual evaluation prompts, refusal behaviour for identity-sensitive queries, and ongoing auditing of model outputs (Mehrabi, Morstatter, Saxena, Lerman, & Galstyan, 2021; Blodgett et al., 2020).

While these approaches can reduce measured bias in some settings, the effects are often partial and depend heavily on the task and evaluation method. A key limitation raised in the literature is that many debiasing techniques reduce surface-level biased outputs without fully removing the underlying associations learned by the model (Bolukbasi, Chang, Zou, Saligrama, & Kalai, 2016; Garg, Schiebinger, Juraf-

sky, & Zou, 2018; Blodgett et al., 2020). For RAG-based educational systems, this implies that post-hoc filtering is not sufficient on its own: bias can already be embedded in the retrieval corpus, meaning that dataset composition and curation become central design concerns rather than afterthoughts.

5.2. Design Considerations for Equitable Deployment

Six high-level design considerations for an LLM-based AI teaching assistant (chatbot) for programming classes are proposed based on the discussed literature synthesis. The chatbot is intended to be used in computing labs at Technological University Dublin, where a human teaching assistant is typically present as well for help and supervision. These considerations aim to help improve experience and address the issue of persistence of women and other underrepresented gender groups in computing higher education.

1. Inclusive Knowledge Architecture. The RAG knowledge base should be carefully curated to reduce representational bias in its retrieved content. This involves ensuring diversity in sources, avoiding gendered framing where unnecessary, and including material that reflects a broad range of contributors and role models in computing, including women's contributions that are often underrepresented in standard course materials. The focus here is on the structure and composition of the retrieval corpus itself, independent of how the system is deployed or embedded institutionally.

2. Stereotype Threat Mitigation by Default. Feedback mechanisms and error-handling language must be designed to minimise stereotype threat for all learners, without relying on gender detection. Growth-oriented framing, normalisation of struggle, and contextual reassurance highlighting diverse role models and normalising capability across groups should be baseline features.

3. Neutral Persona and Tone. Evidence from human–AI interaction research shows that anthropomorphic (assigning human attributes to non-humans) cues, tone, and conversational framing directly shape perceived authority, trust calibration, and engagement in educational contexts (Feine et al., 2020; Waytz, Cacioppo, & Epley, 2009). Overly authoritative or overly human-like personas can lead to automation bias or inappropriate reliance (Parasuraman & Riley, 2000; Dietvorst, Simmons, & Massey, 2015), while under-expressive systems reduce engagement and perceived usefulness (Cassell, 2000; Goetz, Kiesler, & Powers, 2003). The persona should therefore balance neutrality, clarity, and pedagogical transparency to support trust.

4. Communication Functionality. While AI-assisted learning tools have demonstrated some benefits to women in computing education (Leavitt et al., 2022; Maurat, Isip, & Lumabas, 2024), research consistently indicates that women place particular value on peer interaction, with student-to-student interaction being one of the strongest predictor of retention (Barker, McDowell, & Kalahar, 2009) and communal learning environments (Diekmann, Clark, Johnston, Brown, & Steinberg, 2011). Hence, the proposed AI teaching assistant should include built-in peer-to-peer communication alongside an AI-mediated support channel. Students should be able to move fluidly between social and automated forms of assistance, or to seek advice from the AI assistant alone if preferred. Integrating both channels within a single interface reduces friction in switching support modes and helps ensure that lack of confidence does not become a barrier to accessing help.

5. Participatory Co-Design. Women learners, and particularly those from intersectionally marginalised groups, must be active (but not exclusive) co-designers at every stage of the process, from requirements gathering through iterative prototype evaluation.

6. Integration with Intervention Ecosystems. The chatbot should be positioned as one component within coordinated intervention ecosystems, such as the TechMate Toolkit (Berry, McKeever, Murphy, & Delany, 2026), rather than as a standalone solution, ensuring it augments rather than replaces structural, pedagogical, and social support. TechMate (Berry, 2026) was designed to help enhance gender balance in computing education using evidence-based initiatives (actions) organised across four strands: Policy, Pedagogy, Influence & Support, and Promotion & Engagement, consolidating over 400 sources

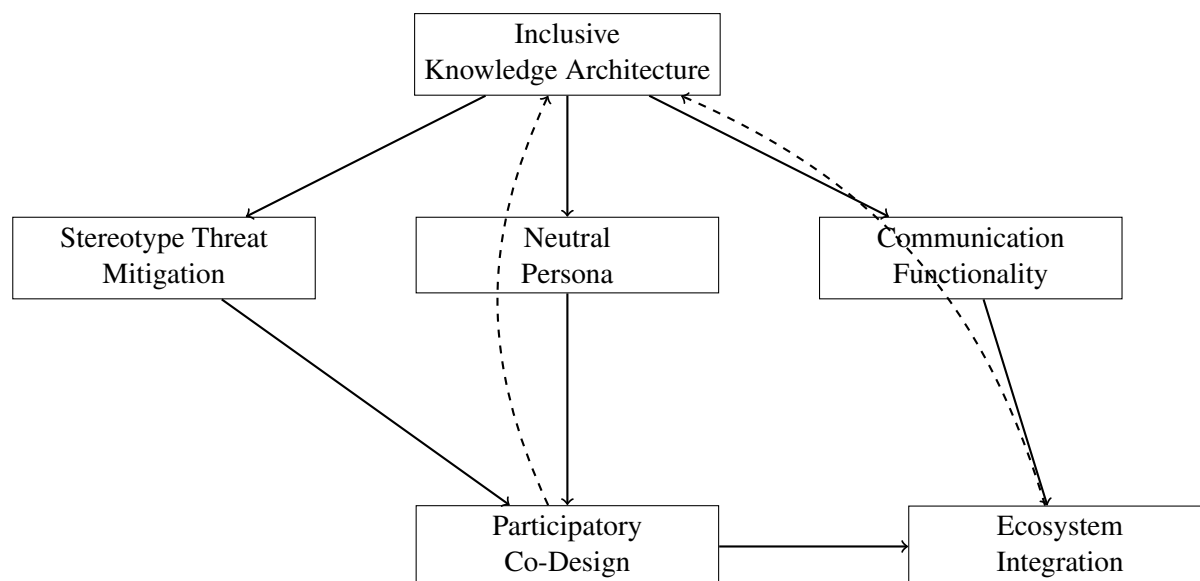


Figure 1 – Graphical representation of the socio-technical model for a gender-responsive RAG-based AI teaching assistant.

into practical, implementation-ready guidance for computing educators (Berry et al., 2026). A chatbot intervention aligns most directly with the Pedagogy and Influence & Support strands, where TechMate identifies peer mentoring, personalised feedback, and inclusive delivery techniques as evidence-based actions with demonstrated impact on women’s retention and belonging in computing (Berry et al., 2026). Deploying the chatbot in conjunction with these existing actions rather than in isolation ensures that technology-mediated support is embedded within a broader institutional commitment to gender equality, addressing the structural and social dimensions of underrepresentation that no single tool can resolve alone (Perez-Felkner et al., 2025; Barker et al., 2009).

This work adopts a socio-technical perspective, which views AI learning systems as interactions between technical architectures, human practices, and institutional contexts rather than isolated algorithmic artefacts (Bostrom & Heinen, 1978; Mumford, 2000; Selbst, Boyd, Friedler, Venkatasubramanian, & Vertesi, 2019; Passi & Barocas, 2020).

Figure 1 presents a socio-technical model in which the RAG knowledge base sits at the centre and shapes the three main design considerations for interaction: stereotype threat mitigation, neutral persona design, and peer learning functionality, while participatory co-design and ecosystem integration act as external but continuous feedback mechanisms that refine and reshape the system over time.

6. Conceptual Evaluation Proposal

To move beyond a design-only contribution, a conceptual evaluation approach for gender-responsive AI teaching assistants in computing education is proposed. The aim is to operationalise the six design considerations into measurable evaluation dimensions for future empirical studies.

Evaluation should occur at three levels. First, system-level evaluation should assess factual accuracy, pedagogical alignment with course material, and consistency of retrieved grounding in RAG-based responses (Liu et al., 2024; Feng et al., 2024). Second, learner-experience evaluation should include confidence, engagement, perceived support, and sense of belonging, which are known to influence persistence in computing education (Dennehy & Dasgupta, 2017; Barker et al., 2009). Third, equity-aware evaluation should disaggregate outcomes across gender and other underrepresented groups, focusing on differences in interaction patterns, help-seeking behaviour, and reliance or avoidance of the system (Martin et al., 2022; Mehrabi et al., 2021). A mixed-methods design is proposed, combining interaction log analysis, pre/post self-report measures, and qualitative student reflections.

7. Conclusion & Next Steps

This paper reviews literature on how AI-assisted educational tools are currently designed, identifies existing bias and mitigation strategies, and situates these issues within the broader challenge of persistence among under-represented gender groups in computing. A socio-technical model is proposed to visualise six design considerations of an AI teaching assistant for computing labs implemented with gender in mind. Future work should empirically evaluate these design considerations in real computing education settings, with attention to differences in confidence, engagement, and persistence across learner groups. This requires moving beyond accuracy-based evaluation toward longitudinal and equity-sensitive assessment, particularly in courses where AI assistants are becoming embedded into everyday learning practice.

It is important to emphasise that the contributions of this paper are currently conceptual rather than empirical. The socio-technical model and the six design considerations are derived from a synthesis of existing literature and have not yet been instantiated in a deployed system. In particular, the proposed guidance on inclusive knowledge architecture and corpus curation has not yet been operationalised into a concrete dataset construction pipeline, and therefore remains at the level of design principles rather than implementation procedures.

In the next step of this work, a prototype LLM- and RAG-based AI teaching assistant will be implemented within a first year programming module (Python or similar) at Technological University Dublin. The system will be deployed during weekly lab sessions with approximately 150 first-year computing students, where it will support debugging, code explanation, and formative feedback tasks alongside existing human teaching assistant support. The study will run over a full teaching semester, allowing tracking of how students interact with the system over time rather than in isolated sessions.

8. References

- Aliyeva, A. (2025). Leveraging rag models to support adaptive learning in computer science at university. In *Proceedings of the panhellenic conference on informatics*. doi: 10.1109/PCI66488.2025.11219854
- Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52–62. doi: 10.61969/jai.1337500
- Barker, L. J., McDowell, C., & Kalahar, K. (2009). Exploring factors that influence computer science introductory course students to persist in the major. *ACM SIGCSE Bulletin*, 41(1), 153–157. doi: 10.1145/1539024.1508982
- Bekkar, A., & Chtouki, Y. (2024). Chatbots in education: A systematic literature review. In *International conference on smart computing and communication*.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610–623). ACM. doi: 10.1145/3442188.3445922
- Berry, A. (2026). *Techmate: Best practice toolkit for gender equality in computing education*. Web-based toolkit. Retrieved from <https://ascnet.ie/techmate/>
- Berry, A., McKeever, S., Murphy, B., & Delany, S. J. (2022). Addressing the “leaky pipeline”: A review and categorisation of actions to recruit and retain women in computing education. In *Edulearn22 proceedings* (p. 9405-9416). IATED.
- Berry, A., McKeever, S., Murphy, B., & Delany, S. J. (2026). Techmate: Translating gender equality research into actionable practice for computing educators. In *Proceedings of the 9th international conference on gender research (icgr 2026)* (Vol. 9). Academic Conferences International Limited. Retrieved from <https://doi.org/10.34190/icgr.9.1.4706> (Published April 25, 2026) doi: 10.34190/icgr.9.1.4706
- Beyer, S. (2014). Why are women underrepresented in computer science? gender differences in stereotypes, self-efficacy, values, and interests and predictors of future cs course-taking and grades.

- Computer Science Education*, 24(2-3), 153–192. doi: 10.1080/08993408.2014.963363
- Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (technology) is power: A critical survey of "bias" in nlp. *ACL*. doi: 10.18653/v1/2020.acl-main.485
- Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *Neurips*.
- Bostrom, R. P., & Heinen, J. S. (1978). Social and technical systems. *Management Science*, 24(8), 857–865.
- Cao, D., Shao, Y., Zhang, L., & Gao, F. (2025). Research on retrieval-augmented architecture design and implications for generative ai-powered question-answering teaching assistant systems. In *Proceedings of the 2025 international conference on advanced computing and intelligent robotics applications*. doi: 10.1109/ACIRA67680.2025.11335016
- Casper, S., Davies, X., Shi, C., Gilbert, T. K., Scheurer, J., Rando, J., ... others (2023). Open problems and fundamental limitations of reinforcement learning from human feedback. *arXiv preprint arXiv:2307.15217*. Retrieved from <https://arxiv.org/abs/2307.15217>
- Cassell, J. (2000). Embodied conversational agents. In *Mit press / aaai fall symposium on socially intelligent agents*.
- Chaturvedi, S., & Chaturvedi, A. (2025). *Who gets the callback? generative AI and gender bias*. arXiv preprint arXiv:2502.XXXXX.
- Chen, Y., Raghuram, V. C., Mattern, J., Mihalcea, R., & Jin, Z. (2025). Causally testing gender bias in llms: A case study on occupational bias. In *Findings of the association for computational linguistics: Naacl 2025* (pp. 4999–5019). doi: 10.18653/v1/2025.findings-naacl.281
- Dakshit, S. (2024). Faculty perspectives on the potential of rag in computer science higher education. In *Proceedings of the conference on information technology education*. doi: 10.1145/3686852.3686864
- de Fine Licht, K. (2024). Generative artificial intelligence in higher education: Why the 'banning approach' to student use is sometimes morally justified. *Philosophy & Technology*, 37(4). doi: 10.1007/s13347-024-00799-9
- Dennehy, T. C., & Dasgupta, N. (2017). Female peer mentors early in college increase women's positive academic experiences and retention in engineering. *Proceedings of the National Academy of Sciences*, 114(23), 5964–5969. doi: 10.1073/pnas.1613117114
- Diekmann, A. B., Clark, E. K., Johnston, A. M., Brown, E. R., & Steinberg, M. (2011). Malleability in communal goals and beliefs influences attraction to STEM careers: Evidence for a goal congruity perspective. *Journal of Personality and Social Psychology*, 101(5), 902–918. doi: 10.1037/a0025199
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. doi: 10.1037/xge0000033
- Elnaffar, R. F., A. A. Z., S. (2025). Teaching with AI: A systematic review of chatbots, generative tools, and tutoring systems in programming education. *International Journal of Learning, Teaching and Educational Research*.
- Eurostat. (2024). ICT specialists in employment.
(https://ec.europa.eu/eurostat/statistics-explained/index.php?title=ICT_specialists_in_employment)
- Feine, J., et al. (2020). Gender bias in chatbot design. In A. Følstad et al. (Eds.), *Chatbot research and design* (Vol. 11970, pp. 79–93). Springer, Cham. doi: 10.1007/978-3-030-39540-7_6
- Feng, T., Liu, S., & Ghosal, D. (2024). Courseassist: Pedagogically appropriate ai tutor for computer science education. In *Proceedings of the sigcse virtual*. doi: 10.1145/3649409.3691094
- Fetrati, H., Chan, G., & Orji, R. (2025). Leveraging generative and rule-based models for persuasive STI education: A multi-chatbot mobile application. In *Proceedings of the 7th acm conference on conversational user interfaces*. Waterloo, ON, Canada: ACM.
- Freeman, J. (2025). *Student generative ai survey 2025* (Tech. Rep.). Oxford: Higher Education

- Policy Institute. Retrieved 2026-08-12, from <https://www.hepi.ac.uk/wp-content/uploads/2025/02/HEPI-Kortext-Student-Generative-AI-Survey-2025.pdf>
- Gallegos, I. O., et al. (2023). Bias and fairness in large language models: A survey. *Computational Linguistics*.
- Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. In *Pnas*. doi: 10.1073/pnas.1720347115
- Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., ... Rienties, B. (2025). The promise and challenges of generative ai in education. *Behaviour & Information Technology*, 44(11), 2518–2544. doi: 10.1080/0144929X.2024.2394886
- Goetz, J., Kiesler, S., & Powers, A. (2003). Fun in learning: The role of embodied conversational agents in education. *Cognitive Systems Research*, 4(4), 409–427. doi: 10.1016/S1389-0417(03)00037-7
- Higher Education Authority. (2026). *Trends in Non-Progression 2016/2017–2023/2024: Course Characteristics*. <https://hea.ie/statistics/data-for-download-and-visualisations/students/progression/hea-progression-and-completion-report-2026/hea-non-progression-course-characteristics-2026/>. (Accessed: 2026-08-12)
- Jamie, P., Hajihashemi, R., & Alipour, S. (2025). Utilizing chatgpt in a data structures and algorithms course: A teaching assistant's perspective. In *Proceedings of the extended abstracts of the chi conference on human factors in computing systems* (pp. 1–7).
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. doi: 10.1016/j.lindif.2023.102274
- Kotek, H., Dockum, R., & Sun, D. (2023). Gender bias and stereotypes in large language models. In *Proceedings of the acm collective intelligence conference (ci '23)*. ACM. doi: 10.1145/3582269.3615599
- Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28(1), 973–1018. doi: 10.1007/s10639-022-11177-3
- Leavitt, C., Corbin, M., Keith, M. J., Schuetzler, R. M., Anderson, G., & Kettles, D. (2022). A chatbot tutor can lessen the gender confidence gap in information systems learning. In *Proceedings of the 55th hawaii international conference on system sciences* (pp. 1–10).
- Lehman, K. J., Newhouse, K. N., Sundar, S., & Sax, L. J. (2023). Nevertheless, they persisted: Factors that promote persistence for women and racially/ethnically minoritized students in undergraduate computing. *Computer Science Education*, 33(2), 260–285.
- Liu, R., Zenke, C., Liu, C., Holmes, A., Thornton, P., & Malan, D. J. (2024). Teaching cs50 with ai: Leveraging generative artificial intelligence in computer science education. In *Proceedings of the technical symposium on computer science education*. doi: 10.1145/3626252.3630938
- Mahajan, R. (2025). Retrieval-augmented generation: The technical foundation of intelligent AI chatbots. *World Journal of Advanced Research and Reviews*.
- Martin, A., Ying, K. M., Rodríguez, F. J., Kahn, C., & Boyer, K. (2022). Intelligent support for all?: A literature review of the (in)equitable design evaluation of adaptive pedagogical systems for cs education. In *Proceedings of the technical symposium on computer science education*. doi: 10.1145/3478431.3499418
- Maurat, J. I. C., Isip, E. V., & Lumabas, A. G. R. (2024). A comparative study of gender differences in the utilization and effectiveness of AI-assisted learning tools in programming among university students. In *Proceedings of the 2024 international conference on artificial intelligence and teacher education*.
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 1–35. doi: 10.1145/3457607
- Mumford, E. (2000). *Socio-technical design: An unfinished science*. Information Systems Institute,

London School of Economics.

- Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science*, 18(10), 879–885. doi: 10.1111/j.1467-9280.2007.01995.x
- Németh, R., Tátrai, A., Szabó, M., & Tamási, j. y. d., Á. (n.d.). Using a rag-enhanced large language model in a virtual teaching assistant role: Experiences from a pilot project in statistics education.
- Ormeño-Arriagada, P., Herrera, B., & Queirolo, F. (2025). Towards inclusive computing: A retrieval-augmented generation prototype for gender equity in computer science curricula. In *Proceedings of the international conference of the chilean computer science society*. doi: 10.1109/SCCC67219.2025.11420736
- Osunde, J. O., Windall, G., Bacon, L., & Mackinnon, L. (2014). Female under-representation in computing education and industry: A survey of issues and interventions. *International Journal of Advanced Computer Science and Applications*, 5(9). doi: 10.14569/IJACSA.2014.051003
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., ... Lowe, R. (2022). Training language models to follow instructions with human feedback. In *Proceedings of the 36th international conference on neural information processing systems*. Red Hook, NY, USA: Curran Associates Inc.
- Pantic, K., & Clarke-Midura, J. (2019). Factors that influence retention of women in the computer science major: A systematic literature review. *Journal of Women and Minorities in Science and Engineering*, 25(2).
- Parasuraman, R., & Riley, V. (2000). Models of automation-induced complacency and automation bias. *Human Factors*, 42(2), 287–301. doi: 10.1518/001872000779698329
- Passi, S., & Barocas, S. (2020). Sociotechnical approaches to ai fairness. In *Proceedings of the aaai/acm conference on ai, ethics, and society (aies)*. ACM.
- Perez-Felkner, L., Erichsen, K., Li, Y., Chen, J., Hu, S., Ramirez Surmeier, L., & Shore, C. (2025). Computing education interventions to increase gender equity from 2000 to 2020: A systematic literature review. *Review of Educational Research*, 95(3), 536–580.
- Pitts, G., Hridi, A. P., & Narayanan, A. B. L. (2025). A survey of llm-based applications in programming education: Balancing automation and human oversight. In *Proceedings of the fourth workshop on bridging human-computer interaction and natural language processing (hci+ nlp)* (pp. 255–262).
- Sadat Shanto, S., Ahmed, Z., & Jony, A. I. (2024). Generative ai for programming education: Can chatgpt facilitate the acquisition of fundamental programming skills for novices? In *Proceedings of the 3rd international conference on computing advancements* (pp. 685–692).
- Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency (fat* / facct)* (pp. 59–68). ACM.
- Smutny, P., & Schreiberova, P. (2026). From rules to language models: a comparative study of chatbot learning assistants. *Frontiers in Education*, 11. Retrieved from <https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2026.1794807/full> doi: 10.3389/feduc.2026.1794807
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797–811.
- Swacha, J., & Gracel, J. (2025). Retrieval-augmented generation (RAG) chatbots for education: A survey of applications. *Applied Sciences*, 15(8), 4234. doi: 10.3390/app15084234
- Szlavi, A., Haugland, H. S., & Jaccheri, L. (2024). Role models as an intervention for gender diversity in computing education. In *Norsk ikt-konferanse for forskning og utdanning*.
- Thompson, L. D., & Wu, Z. (2021). Increasing the enrollment, retention, and graduation of undergraduate women majoring in computing. In *2021 conecd*.
- Wan, Y., Pu, G., Sun, J., Garimella, A., Chang, K. W., & Peng, N. (2023). “kelly is a warm person, joseph is a role model”: Gender biases in llm-generated reference letters. In *Findings of the association for computational linguistics: Emnlp 2023* (pp. 3730–3748).

- Wang, K., & Lawrence, R. (2025). Quantitative evaluation of using large language models and retrieval-augmented generation in computer science education. In *Proceedings of the technical symposium on computer science education*. doi: 10.1145/3641554.3701917
- Waytz, A., Cacioppo, J., & Epley, N. (2009). Making sense by making sentient: Effectance motivation increases anthropomorphism. *Journal of Personality and Social Psychology*, 96(5), 864–877. doi: 10.1037/a0014795
- Yansen, G., & Zukerfeld, M. (2014). Why don't women program? exploring links between gender, technology and software. *Science, Technology and Society*, 19(3), 305–329. doi: 10.1177/0971721814548111
- Yigci, D., Eryilmaz, M., Unal, E., Eryilmaz, N. O., Akcapinar, G., Reisoglu, I., ... Celik, B. (2024). Large language model-based chatbots in higher education. *Advanced Intelligent Systems*. doi: 10.1002/aisy.202400429