ASCILITE 2026

The Future is Connection:

Collaboration, Creativity and Compassion in a Digital Society

Bridging the articulation gap: A multi-agent conversational AI platform for compassionate scaffolding of technical interview skills

Owen Noel Newton Fernando, Dominic Meng Xuan Lai, Siddhanth Manoj, Chin Ann Ong
Nanyang Technological University, Singapore

Timothy Robert Merritt
Aalborg University, Denmark

Gwendoline Choon Lang Quek
National Institute of Education, Singapore

Technical job interviews place computing students under pressure to demonstrate algorithmic and communication skills simultaneously under live observation, producing what we term the articulation gap. Existing preparation pathways provide only partial support: coding platforms ignore communication, coaching is economically inaccessible, and current AI simulators remain reactive. We present AIView, a multi-agent conversational platform that scaffolds integrated interview competencies. A LangGraph orchestrator coordinates specialised agents for question delivery and adaptive guidance alongside verbalisation monitors and an affect-aware tone layer, with assessments anchored in the SOLO taxonomy. An evaluation with 31 students showed high usability (SUS = 73.8), strong trust in AI feedback (M = 3.94), and high recommendation intent (M = 4.29). Participants valued personalised, code-aware feedback and a low-pressure learning environment. By fostering compassionate human-AI collaboration that supports creative problem solving and connection in a digital learning environment, AIView represents a step towards democratising access to career preparation through adaptive scaffolding.

Keywords: multi-agent systems, large language models, scaffolding, cognitive load, compassionate technology, human-centred learning design, employability, SOLO taxonomy, mixed-methods

Introduction

Demand for software engineering professionals continues to expand globally, with computing occupations projected to outpace average employment growth across the next decade. Technical job interviews remain the primary gatekeeping mechanism for entry into the computing profession, requiring candidates to simultaneously solve complex algorithmic problems, articulate design rationale, communicate under conditions of acute uncertainty, and respond adaptively to interviewer prompts (Behroozi et al., 2019, 2020; Ford et al., 2017). The rapid emergence and mainstream adoption of Generative AI (GenAI) coding assistants (e.g., GitHub Copilot, ChatGPT) have fundamentally transformed industry expectations: employers increasingly assume basic syntactic fluency and place decisive evaluation weight on higher-order architectural reasoning, systematic problem decomposition, and verbal communication during live coding (Banh et al., 2025).

Despite this fundamental shift, a substantial preparation gap persists in higher education. University curricula thoroughly prepare students for individual algorithmic problem solving and technical implementation, yet students consistently report feeling severely underprepared for the interpersonal, communicative, and performative demands of technical interviews (Kapoor & Gardner-McCune, 2019). This preparation deficit is particularly pronounced for students from under-represented backgrounds, first-generation university students, and non-native speakers who often lack informal alumni networks, industry coaching, or peer rehearsal opportunities (Ford et al., 2017). The resulting mismatch between internal technical competence and external verbal demonstration constitutes what we term the articulation gap: under dual cognitive load, anxiety, cognitive overload, and hesitation prevent qualified candidates from demonstrating capabilities they genuinely possess (Behroozi et al., 2020).

Current interview preparation pathways address this challenge only partially. Static coding platforms (e.g., LeetCode, HackerRank) train algorithmic fluency in complete silence, wholly ignoring verbalisation, clarification seeking, and real-time design articulation. Peer mock interviews offer interpersonal practice but suffer from inconsistent partner quality, scheduling friction, peer intimidation, and evaluative anxiety. Commercial human coaching provides high-fidelity simulation but is economically prohibitive for most students ($150-$300 per hour), reinforcing socioeconomic inequities in tech hiring. Recent commercial AI interview tools exist, but they predominantly rely on rigid, single-prompt LLM architectures that lack theoretical grounding and remain entirely reactive: they wait passively for the user to speak rather than actively intervening when a candidate falls silent, becomes disoriented, or struggles to structure their problem-solving thoughts.

Recent research demonstrates the immense potential of LLM-based conversational agents for oral assessment, collaborative programming, and instructional scaffolding across educational domains (Dai & Wu, 2025; Wang et al., 2025; Xi et al., 2026; Yusuf et al., 2025). However, limited work has investigated how multi-agent conversational architectures can be orchestrated specifically to scaffold the multi-faceted technical and communicative dimensions of interviews with a compassionate, learner-centred orientation.

This study draws on three foundational theoretical perspectives to inform specific research questions: Cognitive Load Theory (Sweller, 1988) examines whether AIView reduces extraneous cognitive load relative to authentic interviews (RQ3); scaffolding theory (Wood et al., 1976) frames adaptive support mechanisms across varying candidate proficiency tiers (RQ1, RQ4); and the SOLO taxonomy (Biggs & Collis, 1982) underpins evidence-grounded assessment and feedback generation (RQ2):

RQ1. To what extent do computing students perceive AIView as usable, accessible, and acceptable as a technical interview preparation platform, relative to established industry benchmarks?

RQ2. How do students evaluate the trustworthiness, transparency, and adoption potential of AIView's evidence-grounded assessment and feedback mechanisms?

RQ3. Does AIView provide a lower-pressure, more emotionally supportive practise environment than authentic technical interviews while maintaining realistic task engagement?

RQ4. What aspects of the multi-agent scaffolding architecture do students perceive as most valuable or most in need of refinement for verbal and algorithmic skill development?

Theoretical framework and literature review

Cognitive load and the articulation gap

Cognitive Load Theory (Sweller, 1988) posits that working memory is strictly limited in capacity and duration when processing novel information. Sweller distinguishes three forms of cognitive load: intrinsic load (the inherent element interactivity of the task), extraneous load (demands imposed by instructional design or environmental stress), and germane load (mental effort devoted to schema acquisition and automation). In technical job interviews, candidates face exceptionally high intrinsic load from algorithmic problem solving (e.g., managing pointer manipulation, edge cases, time/space complexity) alongside severe extraneous load from live surveillance, whiteboard performance anxiety, unfamiliar social dynamics, and the demanding requirement to "think aloud" concurrently while coding (Behroozi et al., 2019, 2020; Ericsson & Simon, 1993).

When total cognitive load exceeds working memory capacity, performance degrades precipitously – not due to a lack of underlying domain competence, but because working memory is consumed by intrusive performance anxiety and divided attention. This breakdown creates the articulation gap: candidates understand the algorithmic logic internally but cannot simultaneously formulate sentences, code, and respond to verbal cues. Providing adaptive scaffolding that intervenes during moments of silence or confusion can effectively regulate extraneous load, freeing mental bandwidth for structured communication and problem solving (Shin et al., 2025; Wood et al., 1976).

Conversational AI and multi-agent scaffolding

Early educational chatbots relied on deterministic decision trees and pattern-matching scripts, which lacked the contextual adaptability required for nuanced tutoring (Kuhail et al., 2023). While modern Large Language

Models (LLMs) provide sophisticated natural language generation, single-prompt architectures struggle with multi-objective educational tasks where a system must simultaneously inspect code buffers, evaluate algorithmic correctness, monitor candidate affect, deliver pedagogically calibrated hints, and manage pacing. When overloaded with competing objectives, monolithic LLM agents frequently suffer from prompt drift, hallucinated feedback, or premature revelation of full solutions.

Multi-agent architectures address these fundamental limitations by decomposing complex instructional tasks across specialized agents operating in coordinated graphs (Wu et al., 2023). In an educational context, multi-agent frameworks allow dedicated sub-agents to specialize in question framing, conversational hinting, verbalisation monitoring, and assessment. This separation of concerns ensures that instructional support is graduated, structured, and systematically aligned with pedagogical goals.

The AIView system

Design rationale

AIView is designed around three core pedagogical commitments. First, drawing from cognitive load theory (Sweller, 1988), it simulates authentic interview demands while dynamically moderating extraneous load through graduated assistance and active pacing prompts. Second, grounded in constructive alignment (Biggs, 1996), prompts, evaluation rubrics, and feedback directly mirror the holistic criteria used in software engineering hiring: problem comprehension, algorithmic efficiency, code correctness, and communication clarity. Third, consistent with scaffolding theory (Wood et al., 1976), support is adaptive and calibrated to learner progress, gradually fading as competence increases. Rather than functioning as a passive validator, AIView is an active, compassionate partner that intervenes during prolonged silence, confusion, or emotional distress.

System architecture overview

AIView comprises three interlinked subsystems: a multimodal browser interface, a LangGraph multi-agent orchestration framework, and an evidence-grounded assessment pipeline (Figure 1). The client interface integrates a Monaco code editor with syntax highlighting, a live WebRTC audio stream, and an optional webcam video stream for non-invasive affect monitoring. Candidate speech is transcribed in near real-time using OpenAI Whisper, while interviewer responses are synthesised via Azure Neural Text-to-Speech (TTS), producing expressive, human-like voice interaction.

The orchestration layer is built on LangGraph, a stateful graph-based framework developed by LangChain for cyclic, multi-actor LLM applications. Unlike linear chains or AutoGen (Wu et al., 2023), LangGraph explicitly models conversational interviews as directed graphs comprising specialized agent nodes and conditional routing edges. This architecture provides native support for cyclical multi-turn dialogue, shared state persistence, and fine-grained turn classification. All conversational nodes are powered by GPT-4o via Azure OpenAI without fine-tuning; specialized behavior is achieved through system prompts, role decomposition, and selective access to a centralized state schema containing the problem statement, active code buffer, rolling conversation history, and real-time affective signals.

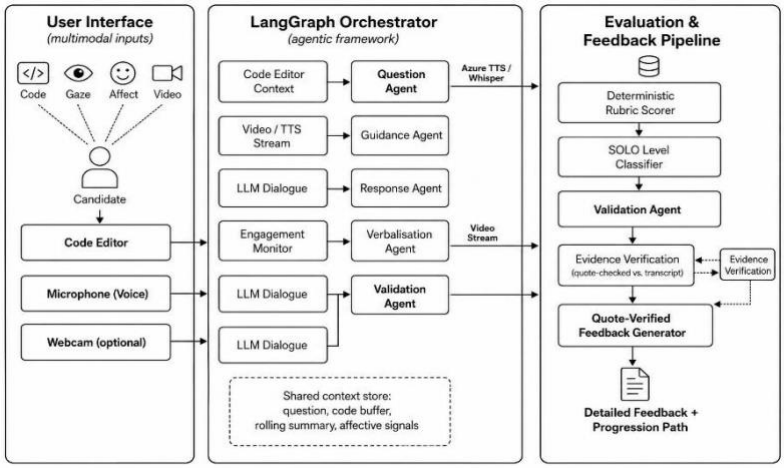


Figure 1. AIView system architecture depicting a LangGraph orchestrator and the feedback pipeline

Multi-agent node architecture

AIView operationalises scaffolding through five specialized agent nodes coordinated within the LangGraph state machine (Table 1). Following established instructional scaffolding taxonomies, agent responsibilities are partitioned between task-oriented support (managing problem progression and hints) and idea-oriented support (clarifying conceptual schemas and encouraging reflection). On each dialogue turn, an orchestrator router node analyzes candidate input against the current interview phase and routes execution to the appropriate specialist agent. A rolling summary node periodically compresses dialogue history, preventing context overflow while preserving essential pedagogical context.

Table 1.
AIView specialist agent nodes and pedagogical functions

Node	Pedagogical type	Function
Question	Task-oriented	Manages structural progression; tests core concepts without premature assistance.
Guidance	Task-oriented scaffolding	Triggered struggle is detected; provides incremental hints to maintain agency while preventing cognitive block.
Response	Idea-oriented	Handles think-aloud validation; context-specific follow-up questions grounded in the candidate's actual approach.
Off-topic	Constraint management	Acts as a guardrail to keep interaction professional, focused, and resilient against prompt injection.

Proactive monitoring and affect-aware adaptation

To address the articulation gap, AIView deploys two background monitoring routines that execute in parallel with the main conversational turn loop. An engagement monitor continuously samples editor events, code changes, and audio input. If a candidate remains silent for more than 15 seconds during active coding, a verbalisation monitor fires a lightweight intervention: "Could you walk me through what you are thinking about right now?" If silence persists beyond 30 seconds alongside repeated compilation errors, control routes directly to the Hint Node, proactively preventing debilitating cognitive overload.

Concurrently, an optional webcam feed analyzes candidate facial expressions at 2 Hz using MediaPipe Face Mesh (Lugaresi et al., 2019; Zhang et al., 2020). The pipeline extracts 468 3D facial landmarks to estimate an empirical affective tension index based on eye-gaze aversion, brow furrowing (Action Units 4 and 7), and head pose jitter. When detected tension exceeds an empirical baseline during multiple execution failures, the tone adaptation layer prepends an empathetic steering directive to the active agent prompt: "The candidate appears stressed. Adopt an encouraging, reassuring tone. Validate their initial conceptual direction before

pointing out the boundary condition." This compassionate framing ensures that technical rigor is balanced with psychological safety.

Evidence-grounded assessment and feedback pipeline

Automated assessment in AI systems faces significant hallucination risks. AIView mitigates these risks through a three-stage verification pipeline. First, an evidence-first extraction step compels the assessment agent to cite verbatim transcript timestamps and exact code line numbers before formulating any evaluative claim. Second, deterministic rubric scoring maps performance across four core competency dimensions: problem comprehension, algorithmic efficiency (Big-O analysis), code correctness (unit test execution), and communication clarity.

Third, conceptual mastery is formally categorized against the SOLO (Structure of Observed Learning Outcomes) taxonomy (Biggs & Collis, 1982), identifying candidate understanding as prestructural, unistructural, multistructural, relational, or extended abstract. Following constructive feedback principles (Erdoğmuş & Kurt, 2024), post-interview reports are presented as explanatory dialogues rather than punitive scorecards, highlighting candidate articulation strengths before providing targeted recommendations for future practice.

Method

Study design

A mixed-methods empirical study was conducted to evaluate AIView's usability, affective support, and scaffolding efficacy. The evaluation combined standardized quantitative instruments, custom validated Likert scales, and qualitative thematic analysis of user experiences. The investigation was approved by the Institutional Review Board (IRB) of the host university and conducted in a dedicated Human-Computer Interaction laboratory in early 2026.

Participants

Participants (*n* = 31; 19 male, 11 female, one non-binary; ages 19-25, M = 21.4, SD = 1.6) were recruited through departmental email lists across undergraduate computing programmes. Eligibility criteria required completion of at least one university-level data structures and algorithms course. The sample reflected diverse preparation backgrounds: 14 participants (45.2%) had prior technical interview experience (internship or full-time recruitment), while 17 participants (54.8%) had never undergone a technical interview. Participation was entirely voluntary, and all participants received a SGD $15 honorarium.

Procedure and instruments

Each evaluation session lasted 60 minutes. After providing informed consent, participants completed a five-minute pre-study questionnaire capturing demographic details and prior interview experience. Participants then received a brief five-minute system onboarding before undertaking an autonomous 30-minute technical interview session on AIView. Two standard interview problems of medium algorithmic difficulty were counterbalanced across participants: "Merge Intervals" (array manipulation, sorting, intervals) and "Valid Parentheses with Wildcards" (stack manipulation, dynamic programming). Participants were instructed to think aloud and explain their rationale as they coded. Proactive verbalisation prompts and webcam affect monitoring were enabled by default, with participants retaining the option to disable camera tracking.

Upon session completion, participants completed a post-study questionnaire consisting of: (a) the standard 10-item System Usability Scale (SUS; Brooke, 1996); (b) nine custom five-point Likert items (1 = Strongly Disagree, 5 = Strongly Agree) assessing assessment transparency, trust in feedback, webcam comfort, recommendation intent, perceived mental effort, anxiety, and stress relative to authentic interviews; and (c) five open-ended qualitative prompts eliciting perceived strengths, frustrations, feedback accuracy, webcam perceptions, and suggestions for improvement. The custom scale items were specifically designed to measure construct dimensions unique to AI-mediated technical interviewing. Internal reliability was confirmed with

Cronbach's α = 0.82. Following the survey, semi-structured interviews (10-15 minutes) were conducted with a purposive subsample of 12 participants representing diverse performance levels.

Quantitative data were analyzed using descriptive statistics (mean, median, standard deviation) and paired t-tests where appropriate. Qualitative survey responses and verbatim interview transcripts were analyzed using reflexive thematic analysis following the six-phase framework of Braun and Clarke (2006). Two researchers independently coded the dataset, establishing strong inter-rater reliability (Cohen's κ = 0.84), with disagreements resolved through collaborative discussion.

Results

System usability (RQ1)

AIView achieved a mean System Usability Scale (SUS) score of 73.8 (Mdn = 75.0, SD = 15.4), well exceeding the widely recognized benchmark of 68.0 and placing the platform in the "Good" usability category (approximately 70th percentile) according to Brooke's (1996) normative distribution. Item-level analysis revealed highest ratings for ease of use (Item 3: M = 4.13, SD = 0.76) and system learnability (Item 7: M = 4.10, SD = 0.79), alongside low perceived inconsistency (Item 6: M = 1.94, SD = 0.81) and minimal perceived complexity (Item 2: M = 2.19, SD = 0.87). Table 2 provides the complete item-level statistics.

Table 2.
SUS item-level descriptive statistics (n = 31). (R) denotes reverse-scored items.

SUS item	M	Mdn	SD
I would like to use AIView frequently	4.03	4.0	0.80
I found AIView unnecessarily complex (R)	2.32	2.0	0.98
I thought AIView was easy to use	4.03	4.0	0.80
I would need technical support (R)	2.06	2.0	1.21
The functions were well integrated	4.29	5.0	0.78
Too much inconsistency (R)	2.39	2.0	1.05
Most people would learn to use AIView quickly	4.35	4.0	0.71
I found AIView cumbersome to use (R)	2.00	2.0	1.24
I felt confident using the system	4.06	4.0	0.73
I needed to learn a lot before getting started (R)	2.48	2.0	1.34
Overall SUS score	**73.8**	**75.0**	**15.4**

Trust, transparency, and adoption intent (RQ2)

Student reception of AIView's feedback and adoption potential was overwhelmingly positive. Recommendation intent received the highest overall agreement (M = 4.29, SD = 0.74), with 87.1% of participants rating it 4 or 5. Trust in AIView's assessment was high (M = 3.94, SD = 0.77), and perceived feedback fairness was similarly favorable (M = 3.84, SD = 0.82). Transparency – understanding how the multi-agent system derived its evaluation – scored M = 3.77 (SD = 0.84). Participants reported strong willingness to

incorporate AIView into their regular interview preparation routines (M = 4.19, SD = 0.75). In contrast, comfort with webcam-based affect monitoring received moderate agreement with substantial variance (M = 3.42, SD = 1.29), reflecting a clear divergence between participants who felt the feature was unobtrusive and supportive (54.8%) and those who expressed heightened camera self-consciousness (32.3%).

Perceived cognitive load (RQ3)

Regarding cognitive load, participants reported moderate overall mental effort during the mock interview (M = 2.90, SD = 0.91 on a five-point scale where three denotes moderate effort). Perceived problem difficulty was rated at M = 3.10 (SD = 0.87). Crucially, participants overwhelmingly agreed that AIView provided a "considerably less stressful practice environment than an authentic interview" (M = 4.23, SD = 0.88), with 83.9% indicating strong agreement (rating 4 or 5). Furthermore, among the 14 participants with prior interview experience, self-reported anxiety during the AIView session was significantly lower than their recalled anxiety during actual technical job interviews (M = 2.26, SD = 0.96 vs. M = 4.10, SD = 0.87, paired $t(13) = 6.42$, $p < .001$, Cohen's $d = 1.98$), confirming a substantial reduction in extraneous evaluative pressure (Koć-Januchta et al., 2022).

Qualitative findings (RQ4)

Reflexive thematic analysis of open-ended responses and interview transcripts revealed two overarching themes. Theme 1: Personalized, Code-Aware Feedback was cited as the most valuable system feature (n = 23/31). Participants highlighted the value of line-specific code annotations, algorithmic time-complexity critiques, and empathetic phrasing. As participant P8 remarked: "The feedback didn't just tell me my solution was $O(n^2)$; it explained how my nested loop could be flattened with a hash set, which felt like a patient senior engineer guiding me." Participants also appreciated the clear breakdown across problem solving, code quality, and communication.

Theme 2: Proactive Verbalisation Prompts Mitigate the Articulation Gap was endorsed by 19 of 31 participants. Candidates noted that timed interventions prevented mental stalling and compelled continuous verbal explanation. P14 explained: "When I get stuck in interviews, I tend to freeze and go completely silent. The system noticing my pause and asking what approach I was considering forced me to externalize my mental model without feeling judged." Constructive critiques centred on speech synthesis latency (n = 8), desiring more aggressive adversarial follow-up questions (n = 6), and requesting customizable controls for webcam monitoring (n = 5).

Discussion

Cognitive load and scaffolded practise

The quantitative and qualitative findings offer compelling empirical validation for our central proposition: candidate technical interview performance can be effectively scaffolded by dynamically modulating cognitive load. Participants rated AIView as significantly less stressful than authentic interviews (M = 4.23) while sustaining authentic cognitive engagement (M = 2.90). By actively intervening during silence and offering timely, conceptual hints rather than full code solutions, AIView's multi-agent orchestrator successfully absorbs extraneous evaluative anxiety, allowing working memory to remain focused on algorithmic problem solving and articulate verbalization.

However, from a methodological standpoint, the cross-sectional, self-report nature of our evaluation precludes definitive causal claims regarding whether reduced anxiety directly translates to superior problem-solving scores in live employment settings. The reported reduction in anxiety reflects perceived state affect during simulated rehearsal. While lower anxiety has well-established theoretical links to enhanced working memory capacity (Sweller, 1988), longitudinal intervention studies tracking pre-to-post performance in authentic hiring pipelines are necessary to establish direct transfer of communicative competence.

The tool-logic challenge

A salient qualitative finding relates to what we define as the tool-logic challenge. Despite natural neural voice synthesis and proactive conversational cues, some participants initially approached AIView with a transactional mental model, treating the platform as a code compiler or passive question search engine rather than an interactive conversational interviewer. This was evidenced by concise, command-driven inputs ("Check my code") instead of professional communicative articulation ("I am choosing a hash map here to achieve linear time complexity").

Grounded in our empirical observations, participants who engaged comfortably with the webcam affect monitoring reported substantially higher communicative presence and self-awareness (Theme 2). In contrast, participants who felt camera discomfort tended to retreat to minimal text interactions. Designing structured onboarding modules that explicitly model reciprocal conversational expectations is vital for helping candidates transition from solitary coding habits to professional communicative interview dynamics.

Implications for equity

The technical interview articulation gap carries profound implications for educational equity and social mobility. Traditional tech hiring processes disproportionately favor candidates with access to private industry coaching, professional alumni networks, and institutional rehearsal cultures (Ford et al., 2017). Candidates from under-represented backgrounds or non-traditional pathways frequently possess solid technical competence but falter under the performative communicative demands of live interviews. By providing free, accessible, affect-aware multi-agent interview scaffolding, platforms like AIView offer a scalable pathway toward democratizing career preparation and broadening participation in computing (Heymann et al., 2024).

Limitations

Several methodological and technological limitations must be noted. First, the evaluation was conducted with 31 computing students from a single research university; while diverse in interview experience, replication across different institutional contexts and international cohorts is warranted. Second, the study evaluated a single 30-minute practice session; long-term skill retention and behavioral transfer across multiple practice intervals remain unmeasured. Third, affective tension was inferred from 2D webcam video via facial landmark estimation without physiological validation (e.g., galvanic skin response or heart-rate variability). Fourth, the coding tasks focused on core data structures and algorithms, excluding systems design or collaborative pair-programming scenarios. Finally, self-selection bias may have attracted students predisposed to technology-mediated learning tools.

Conclusion

AIView demonstrates that multi-agent conversational AI can effectively bridge the articulation gap by scaffolding both technical problem-solving and verbal communication skills. By coordinating specialized pedagogical agents through LangGraph, integrating proactive verbalisation monitors, adapting tone compassionately to candidate affect, and grounding assessments in the SOLO taxonomy, the platform achieves high usability (SUS = 73.8) while substantially alleviating performance anxiety. In alignment with the ASCILITE 2026 theme – connection, collaboration, and compassion in a digital society – AIView illustrates how empathetic AI systems can empower learners, democratize interview readiness, and foster genuine professional growth.

References

Banh, L., Holldack, F., & Strobel, G. (2025). Copiloting the future: How generative AI transforms software engineering. *Information and Software Technology, 183*, 107751. https://doi.org/10.1016/j.infsof.2025.107751

Behroozi, M., Parnin, C., & Barik, T. (2019). Hiring is broken: What do developers say about technical interviews? In *Proceedings of the IEEE Symposium on Visual Languages and Human-Centric Computing* (pp. 1-9). IEEE. https://doi.org/10.1109/VLHCC.2019.8818836

Behroozi, M., Shirolkar, S., Barik, T., & Parnin, C. (2020). Debugging hiring: What went right and what went wrong in the technical interview process. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Software Engineering in Society* (pp. 71-80). ACM. https://doi.org/10.1145/3377815.338137

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

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. Academic Press.

Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, I. L. McClelland, & B. Weerdmeester (Eds.), *Usability evaluation in industry* (pp. 189-194). Taylor & Francis.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Dai, L., & Wu, F. (2025). An AI-powered conversational system for college students learning English as a second language. *Education and Information Technologies, 30*, 23393-23417. https://doi.org/10.1007/s10639-025-13640-3

Erdoğmuş, Y. K., & Kurt, A. A. (2024). Digital game-based learning: Pedagogical agent and feedback types on achievement, flow experience, and cognitive load. *Education and Information Technologies*, *29*, 12943–12968. https://doi.org/10.1007/s10639-023-12368-2

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). MIT Press.

Ford, D., Barik, T., Rand-Pickett, L., & Parnin, C. (2017). The tech-talk balance: What technical interviewers expect from technical candidates. In *Proceedings of the 10th International Workshop on Cooperative and Human Aspects of Software Engineering* (pp. 43-48). IEEE. https://doi.org/10.1109/CHASE.2017.8

Heymann, P., Hukema, M., van Rosmalen, P., & Beausaert, S. (2024). Towards design principles for an online learning platform providing reflective practises for developing employability competences. *Education and Information Technologies*, 29, 17211–17234. https://doi.org/10.1007/s10639-024-12530-4

Kapoor, A., & Gardner-McCune, C. (2019). Understanding CS undergraduate students' professional identity through the lens of their technical interview experiences. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (pp. 710-716). ACM. https://doi.org/10.1145/3304221.3319764

Koć-Januchta, M. M., Schönborn, K. J., Roehrig, C., Chaudhri, V. K., Tibell, L. A. E., & Heller, H. C. (2022). Connecting concepts helps put main ideas together: Cognitive load and usability in learning biology with an AI-enriched textbook. *International Journal of Educational Technology in Higher Education, 19*, 11. https://doi.org/10.1186/s41239-021-00317-3

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies, 28*, 973-1018. https://doi.org/10.1007/s10639-022-11177-3

Shin, Y., Jung, J., & Choi, S. & Jung, B. (2025). The influence of scaffolding for computational thinking on cognitive load and problem-solving skills in collaborative programming. *Education and Information Technologies*. 30(1):583-606. DOI: https://doi.org/10.1007/s10639-024-13104-0

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M. G., Lee, J., Chang, W.-T., Hua, W., Georg, M., & Grundmann, M. (2019). MediaPipe: A framework for building perception pipelines. *arXiv.* https://doi.org/10.48550/arXiv.1906.08172

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285. https://doi.org/10.1207/s15516709cog1202_4

Wang, H., Wang, C., Chen, Z., Liu, F., Bao, C. & Xu, X. (2025). Impact of AI-agent-supported collaborative learning on the learning outcomes of university programming courses. *Education and Information Technologies*, 30, 17717–17749. https://doi.org/10.1007/s10639-025-13487-8

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89-100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Wu, Q., Bansal, G., Zhang, J., Wu, Y., Li, B., Zhu, E., Jiang, L., Zhang, X., Zhang, S., Liu, J., Awadallah, A. H., White, R. W., Burger, D., & Wang, C. (2023). AutoGen: Enabling next-gen LLM applications via multi-agent conversation. *arXiv*. https://doi.org/10.48550/arXiv.2308.08155

Xi, L., Zhang, Y., & Wang, Q. (2026). Investigating the effects of an LLM-based Socratic conversational agent on students' academic performance and reflective thinking in higher education. *Computers & Education, 241*, 105494. https://doi.org/10.1016/j.compedu.2025.105494

Yusuf, H., Money, A., & Daylamani-Zad, D. (2025). Towards reducing teacher burden in performance-based assessments using AIvaluate: An emotionally intelligent LLM-augmented pedagogical AI conversational agent. *Education and Information Technologies, 30*, 24649-24693. https://doi.org/10.1007/s10639-025-13755-7

Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C.-L., & Grundmann, M. (2020). MediaPipe Hands: On-device real-time hand tracking. *arXiv*. https://doi.org/10.48550/arXiv.2006.10214

Fernando, O. N. N., Lai, D. M. X., Manoj, S., Ong, C. A., Merritt, T. R., & Quek, C. L. G. (2026). Bridging the articulation gap: A multi-agent conversational AI platform for compassionate scaffolding of technical interview skills. In Cochrane, T., Narayan, V., Poole, T., Lai, P., Lake, P., Moore, O., & Li, L. (Eds.), *The Future is Connection: Collaboration, Creativity and Compassion in a Digital Society. Proceedings ASCILITE 2026. Brisbane* (pp. 159-168). https://doi.org/10.65106/apubs.2026.3135

Note: All published papers are refereed, having undergone a double-blind peer-review process.

The author(s) assign a Creative Commons by attribution license enabling others to distribute, remix, tweak, and build upon their work, even commercially, as long as credit is given to the author(s) for the original creation.

© Fernando, O. N. N., Lai, D. M. X., Manoj, S., Ong, C. A., Merritt, T. R., & Quek, C. L. G. 2026