

GENERATIVE ARTIFICIAL INTELLIGENCE IN VOCATIONAL UNDERGRADUATE FINANCE AND ACCOUNTING EDUCATION: APPLICATIONS, RISKS, AND FUTURE DIRECTIONS

BiJie Li*, HaoXuan Li

School of Finance and Economics, Hainan Vocational University of Science and Technology, Haikou 571126, Hainan, China.

**Corresponding Author: BiJie Li*

Abstract: Generative artificial intelligence is entering finance and accounting classrooms before curricula, assessment systems, and institutional controls have fully adapted. This narrative review explains the technical principles of large language models and examines their use in vocational undergraduate finance and accounting education. Potential applications include dialogic tutoring, formative feedback, text simplification, case simulation, spreadsheet and programming support, financial analysis, professional communication, and course design. These functions can make practice more accessible and provide rapid opportunities for comparison and revision. Their educational value, however, depends on how the task is structured. Fluent output may contain fabricated facts, obsolete rules, incorrect calculations, biased assumptions, or unsupported recommendations. Unrestricted assistance may also improve immediate task performance without developing the independent knowledge and professional judgment that students need. Additional concerns include academic misconduct, confidential-data exposure, unequal access, unreliable detection, and weak accountability. A responsible integration model should therefore separate AI-free foundations, guided AI assistance, and AI-critical professional tasks. It should require authoritative-source verification, calculation checks, process records, disclosure, reflection, and oral or practical defense. Future systems may combine domain-controlled retrieval, multimodal simulations, and auditable learning records, but vocational programmes need direct longitudinal and multi-institutional evidence. Generative AI should be treated as a supervised learning partner and an object of professional evaluation, not as an authoritative answer engine.

Keywords: Large language models; Accounting education; Finance education; Vocational undergraduate education; Authentic assessment

1 INTRODUCTION

Generative artificial intelligence (GenAI) has changed the cost and speed of producing explanations, examples, computer code, summaries, and conversational feedback. In higher education, large language models can support interaction and differentiated assistance, but their usefulness is inseparable from factual error, bias, overreliance, and the need for critical verification [1]. These tensions are unusually important in finance and accounting. A grammatically polished answer may still apply the wrong reporting rule, omit an assumption, mishandle a discount rate, invent a citation, or expose confidential information.

The issue is not simply whether students should be allowed to use a chatbot. Accounting education scholars argue that GenAI presents both a threat to established teaching practices and an opportunity to move beyond narrow technical reproduction toward ethical reasoning, contextual interpretation, and human judgment [2]. In finance education, students and educators report useful applications of AI tools in learning and financial modelling, although much of the evidence concerns perceptions and experiences rather than measured learning gains [3]. A 2025 survey at one United States university similarly found frequent student use of AI for difficult accounting concepts, but its self-reported benefits cannot establish that AI caused higher achievement [4].

Vocational undergraduate programmes require a particularly careful response because they connect bachelor's-level knowledge with occupational performance. The OECD reports that AI is beginning to affect vocational curricula, qualifications, occupational standards, career guidance, and programme delivery [5]. For finance and accounting programmes, this makes workplace relevance essential, but it also makes independent competence, confidentiality, evidence, and accountability non-negotiable.

This paper presents a selective narrative review of GenAI's technical principles, educational applications, risks, responsible integration, and future directions. It applies international evidence from higher education, accounting, finance, and vocational education to the vocational undergraduate context. Its central argument is that GenAI can improve access to practice and feedback only when students must first think, then use the tool within explicit limits, and finally verify and defend the resulting work.

2 TECHNICAL PRINCIPLES AND THEIR EDUCATIONAL MEANING

2.1 Transformer-Based Generation and Instruction Alignment

Most contemporary large language models are built on the Transformer architecture, which uses self-attention to represent relationships among tokens in a sequence and permits highly parallel computation [6]. During pre-training, a generative model learns statistical regularities from large text and code corpora by predicting tokens from context. At use time, it repeatedly selects a probable continuation conditioned on the prompt and preceding output. This process can produce coherent explanations and plausible professional language, but it does not turn the model into an accounting database, a financial calculator, or a source of professional authority.

Post-training makes the interface more useful. Instruction tuning teaches a model to respond to task directions, while reinforcement learning from human feedback uses ranked responses to bring outputs closer to human preferences. The InstructGPT study showed that these techniques improved instruction following, truthfulness, and perceived helpfulness on its evaluation distribution, yet the authors also reported that aligned models continued to make simple mistakes [7]. A compliant response is therefore not necessarily a correct one. Prompt quality can influence relevance and format, but no prompt can guarantee conformity with a particular reporting standard, tax rule, market date, or institutional policy.

2.2 Grounding, Retrieval, and Verification

Retrieval-augmented generation (RAG) combines a language model with an external document index. The system retrieves potentially relevant passages and conditions generation on them, making knowledge easier to update and, in principle, easier to trace [8]. In a finance or accounting course, a controlled RAG system could search approved textbooks, instructor notes, accounting standards, regulatory publications, and worked cases rather than relying only on knowledge encoded during pre-training.

Grounding reduces some risks but does not remove them. The retriever may select an irrelevant passage, the collection may contain an obsolete version, and the model may misstate the retrieved evidence. Hallucination research distinguishes several forms of generated content that are unfaithful to the source or unsupported by available information [9]. Educational systems should therefore display source identity, jurisdiction, effective date, and retrieved passage, while students should check the original document. Calculations require an independent spreadsheet, calculator, or tested program. A citation supplied by the model should be treated as a search lead until its bibliographic record and substantive claim have been verified.

Table 1 links common technical functions with their educational consequences.

Table 1 GenAI Functions and Instructional Controls in Finance and Accounting Education

Function	Suitable educational role	Main failure mode	Minimum control
Conversational generation	Explain a concept, ask diagnostic questions, propose examples	Fluent but incorrect or overgeneralized answer	Require a student explanation and comparison with an approved source
Retrieval-augmented generation	Locate relevant standards, policies, or course materials	Wrong document, version, jurisdiction, or passage	Display provenance and verify against the original document
Code and formula generation	Draft spreadsheet formulas, SQL, Python, or calculation steps	Silent logic, unit, sign, or data-range error	Test with known cases and reconcile outputs independently
Scenario generation	Simulate clients, analysts, auditors, managers, or regulators	Stereotyped, unrealistic, or internally inconsistent case	Instructor validates the facts, roles, and scoring rubric
Text generation and revision	Draft or critique memos, reports, and presentations	Ghostwriting, fabricated citations, loss of student voice	Preserve drafts, disclose use, and require an oral or practical defense

3 EDUCATIONAL APPLICATIONS

3.1 Dialogic Tutoring and Formative Feedback

GenAI can provide an always-available dialogue in which a student asks for another explanation, requests a worked example, or receives a hint matched to a current difficulty. This can be useful in large or mixed-ability classes where individual instructor time is limited. Suitable tasks include explaining debit and credit effects, comparing valuation methods, identifying the assumptions in a ratio, interpreting the time value of money, or asking a student to justify each step in a solution.

The instructional design matters more than immediate assisted performance. A large field experiment in high-school mathematics found that access to a general GPT interface improved practice performance but reduced subsequent unaided performance, whereas a tutor constrained to provide teacher-designed hints largely mitigated the harm [10]. The population and subject differ from vocational undergraduate finance and accounting, so the result should not be treated as direct programme evidence. It nevertheless provides a strong warning against answer-first tutoring. A better sequence is independent attempt, targeted hint, revised solution, source or calculation check, and a short explanation of what changed.

Accounting classroom evidence points in the same direction. In a financial accounting activity, students answered first, reviewed an AI response, and then revised and evaluated their own work; the experience indicated that prior knowledge affected how effectively students could prompt and judge the tool [11]. In an accounting data analytics course involving 312 students, ChatGPT expertise was positively associated with course performance, especially among students with

lower prior grade-point averages, and students valued its accessible, low-pressure support [12]. Because this was an observational study in one course, the association does not prove a causal learning benefit. It does suggest that programmes should teach help-seeking and verification explicitly instead of assuming students will acquire these practices informally.

3.2 Accessible Materials and Feedback for Revision

Teachers can use GenAI to produce alternative explanations, adjust reading difficulty, create additional practice questions, and draft feedback aligned with a rubric. This is useful when students enter vocational programmes with different mathematical, linguistic, or computing backgrounds. An instructor might ask the model to simplify a dense explanation of revenue recognition, create parallel examples with different data, or identify where a student's memo lacks evidence.

Simplification must preserve meaning. A study of ChatGPT-modified financial accounting examination texts found increased perceived comprehension and reduced confusion and frustration, but it did not document an overall positive effect on performance [13]. Readability and learning are not interchangeable. Teachers should compare simplified text with the original technical requirement, retain defined terms, and review every generated question and answer key before use. Automated feedback should direct attention to reasoning and evidence rather than merely provide a polished replacement answer.

3.3 Cases, Simulations, and Professional Judgment

Vocational curricula can use GenAI to simulate workplace conversations and incomplete business situations. The model can assume the role of a client explaining an unusual transaction, a credit officer questioning cash-flow assumptions, an audit manager challenging evidence, or a board member requesting a concise analysis. Students can then identify missing information, ask follow-up questions, prepare a recommendation, and defend it. Instructors can also seed deliberate errors or conflicting evidence so that evaluation, not acceptance, becomes the learning objective.

Accounting-case research supports this critical-use design. Evidence from eight case-based assignments suggests that current systems are not uniformly competent across accounting work: they cope more reliably with interpretive and judgment-oriented prompts than with tasks that require constructing formal accounting outputs or operating specialised software [14]. The study assessed model capability, not student learning, and newer models may behave differently. Its main pedagogical lesson is durable: performance varies by task. Students should treat an AI-generated solution as work prepared by an imperfect junior assistant and document which claims, entries, assumptions, and calculations they accepted, corrected, or rejected.

Quantitative finance requires additional caution. A study that tested ChatGPT, Bard, and Bing AI on 20 undergraduate quantitative-finance questions reported low scores and errors involving formulas, logarithms, iterative calculations, and arithmetic [15]. The models and test date limit its relevance to later systems, but the error categories remain useful for instruction. AI may explain a formula or draft code, while students must select the appropriate financial method, confirm units and timing, test boundary cases, and reproduce the result with an independent tool.

3.4 Data Analytics, Coding, and Professional Communication

GenAI can lower the initial barrier to spreadsheet formulas, SQL queries, Python scripts, visualizations, and written interpretations. In finance education, reported uses include supporting financial modelling and time-series programming [3]. In accounting data analytics, it can help students understand error messages, translate a business question into code, or compare alternative visualizations [12]. The educational goal should remain disciplined data reasoning: defining the question, inspecting data quality, choosing an appropriate method, testing the output, and explaining its business meaning.

Only synthetic, public, or institutionally approved data should be entered into external systems. Students should retain the prompt, generated code, test cases, corrections, and final interpretation. For professional communication, the model can produce contrasting versions of an audit finding, credit memorandum, variance explanation, or management presentation. Students can evaluate tone, materiality, evidence, and audience suitability, then rewrite the document in their own accountable voice.

4 RISKS AND CHALLENGES

4.1 Hallucination, Currency, and Numerical Reliability

GenAI optimizes plausible continuation, not compliance with professional standards. A response can combine real terminology with a fabricated paragraph number or cite an authentic standard for a claim it does not contain. Rules also change across dates and jurisdictions. Every standards-based task should therefore specify the reporting framework, jurisdiction, effective date, and authoritative source hierarchy. The final submission should include the exact source and locator used for verification.

Numerical answers require a second evidential chain. Students should expose formulas, assumptions, units, signs, rounding, time periods, and data ranges. A correct final number reached through invalid reasoning does not demonstrate

competence. Conversely, a student who detects and explains a model error may show more learning than one who submits an unexamined correct result.

4.2 Privacy, Bias, and Unequal Access

Finance and accounting exercises often contain information that would be sensitive in practice. Uploading client files, payroll records, tax information, unpublished results, personal identifiers, or assessment data to an unapproved service can breach confidentiality even when the subsequent output appears harmless. UNESCO's guidance calls for human-centred educational use, protection of personal data, ethical and pedagogical validation, inclusion, and preservation of human agency [16]. Institutions should provide approved tools, data-processing terms, retention controls, role-based access, and safe practice datasets.

Bias can enter through training data, prompts, retrieval collections, or teacher-designed cases. A model may reproduce stereotypes when describing borrowers, employees, entrepreneurs, or regions, or it may present dominant-market assumptions as universal. Students should compare outputs across neutral counterfactual cases and ask whether irrelevant personal attributes altered the recommendation. Equity also concerns access: paid systems, faster devices, stronger English proficiency, and prior computing experience can produce unequal assistance. Core learning outcomes must remain achievable with institutionally supplied resources.

4.3 Academic Integrity and Assessment Validity

Unclear rules create avoidable misconduct. A survey of 279 accounting students at Jordanian public universities linked pressure, opportunity, and rationalization with ChatGPT misuse under the academic misconduct triangle framework. Its self-reported, country-specific evidence should not be generalized without qualification, but it shows why a policy based only on prohibition is incomplete. Students need task-level rules explaining whether AI is prohibited, permitted for specified functions, or required and evaluated.

Academic-integrity scholarship recommends clear expectations, student education, and assessment change in response to ChatGPT [17]. Detection software should not be the main control. Model capability changes, generated text can be edited, and false accusations carry serious consequences. Stronger assessment combines individualized data, staged submissions, in-class or supervised components, prompt and revision records, source verification, and short oral questioning. These features make reasoning visible and reduce the value of submitting an answer that the student cannot explain.

4.4 Overreliance, Deskilling, and Accountability

Students cannot effectively audit an answer in a domain they have never learned. If AI completes early practice in double-entry logic, reconciliation, formula selection, or code construction, it may remove the productive effort through which mental models develop. Programmes should preserve AI-free foundation tasks and periodically test unaided performance. Assistance can increase only after students demonstrate the prerequisite skill.

Accountability must also remain identifiable. The student is responsible for submitted work, and the instructor remains responsible for teaching materials, feedback, and assessment decisions. A vendor's model update can change outputs without warning, so prompts alone do not ensure reproducibility. Important activities should record the model or system, access date, input materials, relevant settings, output, verification steps, and human decision.

5 A RESPONSIBLE INTEGRATION MODEL

5.1 Three Levels of Permitted Use

A practical curriculum can classify tasks into three levels. Level 1 is AI-free foundation work, used for core terminology, accounting logic, numerical fluency, and independent explanation. Level 2 is guided AI assistance, in which the tool may provide hints, examples, debugging support, or feedback after an initial attempt. Level 3 is AI-critical professional work, where students are expected to use the system, audit its output, consult authoritative evidence, and defend a final judgment.

This progression treats AI literacy as more than prompt technique. Research conceptualizes AI literacy as including knowledge and understanding, use and application, evaluation and creation, and ethical awareness [18]. Finance and accounting programmes should add domain-specific elements: source hierarchy, professional skepticism, data governance, calculation validation, documentation, materiality, and responsibility for decisions.

5.2 Assessment and Evidence of Learning

Assessment should capture both the product and the learning process. A useful AI-assisted submission can contain the student's initial analysis, allowed prompts, selected outputs, an error-and-verification log, authoritative sources, calculation files, the revised conclusion, and a reflection on where human judgment changed the result. Oral defense, live modification of assumptions, or a short unaided transfer problem can test whether knowledge remains available when the tool is removed.

An analysis of 18 assessment tasks across six core undergraduate accounting courses found that ChatGPT was more capable on lower- and middle-order activities than on contextualized standards application, complex calculation, linked questions, ambiguity, and higher-order judgment [19]. This supports assessment designs that move beyond information retrieval and isolated answers. High-stakes tasks should combine context, multiple evidence types, sequential decisions, and accountability for uncertainty rather than rely on complexity merely to defeat a current model.

5.3 Programme Governance and Faculty Capacity

Responsible integration needs programme-level consistency. Each course should publish an AI-use matrix, disclosure format, privacy rule, approved source list, and consequences for misuse. Departments should review tools for data handling, accessibility, security, cost, version change, and pedagogical fit. Teachers need time to test activities and learn enough about model behaviour to identify failure modes. Collaboration with employers can provide realistic but de-identified workflows, while accounting and finance faculty should retain control over learning outcomes and assessment judgments.

Table 2 summarizes the progression.

Table 2 Progressive Integration across a Vocational Undergraduate Programme

Level	Typical task	Permitted AI role	Evidence required from the student
1. AI-free foundation	Journal entries, ratio logic, formula derivation, basic reconciliation	None during the assessed task	Independent work and explanation
2. Guided assistance	Concept review, code debugging, practice cases, draft feedback	Hints or critique after an initial attempt	Initial attempt, prompt/output record, revisions, verification
3. AI-critical professional work	Financial analysis, audit case, credit memo, dashboard, client communication	Required as a tool whose output must be evaluated	Source trail, calculation tests, error log, disclosure, final defense

6 FUTURE DIRECTIONS

6.1 Domain-Grounded and Auditable Learning Systems

Future educational systems are likely to use RAG over licensed, version-controlled collections of standards, regulations, textbooks, and course cases. Effective systems should show the retrieved passage, date, jurisdiction, and source version rather than provide an untraceable answer. They should also preserve prompts, outputs, feedback, and student revisions so that teachers can examine the learning process without treating constant surveillance as pedagogy.

Multimodal and agent-based simulations could place students in controlled ERP, audit, banking, or investment environments. A student might inspect documents, question a simulated client, write code, reconcile records, and present a recommendation while the system changes assumptions. Such environments can increase practice opportunities, but instructor-validated facts and reproducible scoring will be essential.

6.2 Evidence, Competencies, and Inclusion

Research should move beyond adoption intention and self-reported usefulness. Vocational undergraduate programmes need multi-institutional and longitudinal studies that compare unaided learning, guided assistance, and unrestricted use; measure transfer after AI access is removed; and examine professional judgment, data skills, communication, ethical reasoning, and retention. Evidence should also be reported by prior attainment, language, disability, access conditions, and programme type.

Curricula will need continuous review as model capabilities and finance occupations change [5]. The durable outcome is not proficiency with one product. It is the ability to frame a business problem, choose and supervise appropriate tools, evaluate evidence, protect data, communicate uncertainty, and accept responsibility for the decision.

7 REVIEW LIMITATIONS

The evidence was synthesised selectively and narratively; no systematic-review protocol or quantitative pooling procedure was used. Direct evidence from vocational undergraduate finance and accounting programmes remains limited. Many cited studies examine one institution, one course, early model versions, self-reported perceptions, or model performance rather than student learning. The proposed integration model therefore requires contextual testing rather than being treated as a validated universal design.

8 CONCLUSION

Generative AI can expand the amount and variety of practice available in vocational undergraduate finance and accounting education. It can explain concepts, generate cases, simplify texts, assist coding, simulate professional dialogue, and critique drafts. Yet the same interface can conceal unsupported facts, obsolete rules, numerical mistakes, biased assumptions, confidentiality breaches, and work the student cannot independently perform.

The appropriate response is neither unrestricted adoption nor a general ban. Programmes should preserve independent foundations, introduce guided assistance after an initial attempt, and culminate in professional tasks where students

must audit AI output. Source verification, calculation tests, process evidence, disclosure, privacy controls, reflective revision, and oral or practical defense should be normal parts of assessment. Teachers and institutions remain responsible for tool selection and instructional design, while students remain responsible for the claims and decisions they submit. Under these conditions, GenAI can support practice-oriented learning while strengthening, rather than displacing, the knowledge, skepticism, and judgment expected of finance and accounting professionals.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

- [1] Kasneci E, Sessler K, Küchemann S, et al. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 2023, 103: 102274. DOI: 10.1016/j.lindif.2023.102274.
- [2] Ballantine J, Boyce G, Stoner G. A critical review of AI in accounting education: Threat and opportunity. *Critical Perspectives on Accounting*, 2024, 99: 102711. DOI: 10.1016/j.cpa.2024.102711.
- [3] Córdova P, Grájeda A, Córdova J P, et al. Leveraging AI tools in finance education: Exploring student perceptions, emotional reactions and educator experiences. *Cogent Education*, 2024, 11(1): 2431885. DOI: 10.1080/2331186X.2024.2431885.
- [4] Zhou A, Luo Y. Exploring the impact of generative AI on student learning in accounting. *Journal of Accounting Education*, 2025, 72: 100982. DOI: 10.1016/j.jaccedu.2025.100982.
- [5] Organisation for Economic Co-operation and Development. *Developing Vocational Education and Training with Artificial Intelligence*. Paris: OECD Publishing, 2026. DOI: 10.1787/e9f76b4e-en.
- [6] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017, 30: 5998-6008. <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>.
- [7] Ouyang L, Wu J, Jiang X, et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 2022, 35: 27730-27744. DOI: 10.52202/068431-2011.
- [8] Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 2020, 33: 9459-9474. <https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html>.
- [9] Ji Z, Lee N, Frieske R, et al. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 2023, 55(12): 1-38. DOI: 10.1145/3571730.
- [10] Bastani H, Bastani O, Sungu A, et al. Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 2025, 122(26): e2422633122. DOI: 10.1073/pnas.2422633122.
- [11] Calderon T G, Gao L, Cardoso R L. Generative artificial intelligence in the classroom: A financial accounting experience. *Advances in Accounting Education: Teaching and Curriculum Innovations*, 2023, 27: 125-144. DOI: 10.1108/S1085-462220230000027006.
- [12] Blondeel E, Bullock T, Gaskin J, et al. The effects of generative artificial intelligence (GenAI) on learning in an accounting data analytics course. *Journal of Accounting Education*, 2025, 72: 100987. DOI: 10.1016/j.jaccedu.2025.100987.
- [13] Voshaar J, Weeks J O, Plate B J, et al. Tackling professorial expert bias: The role of ChatGPT in simplifying financial accounting exam texts. *Issues in Accounting Education*, 2025, 40(1): 93-123. DOI: 10.2308/ISSUES-2023-091.
- [14] Cheng X, Dunn R, Holt T, et al. Artificial intelligence's capabilities, limitations, and impact on accounting education: Investigating ChatGPT's performance on educational accounting cases. *Issues in Accounting Education*, 2024, 39(2): 23-47. DOI: 10.2308/ISSUES-2023-032.
- [15] Malladi R K. Emerging frontiers: Exploring the impact of generative AI platforms on university quantitative finance examinations. *Journal of Financial Education*, 2024, 50(2): 45-100. <https://jfedweb.org/issues/JFEV5002.pdf>.
- [16] Miao F, Holmes W. *Guidance for generative AI in education and research*. Paris: UNESCO, 2023. DOI: 10.54675/EWZM9535.
- [17] Cotton D R E, Cotton P A, Shipway J R. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 2024, 61(2): 228-239. DOI: 10.1080/14703297.2023.2190148.
- [18] Ng D T K, Leung J K L, Chu S K W, et al. Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2021, 2: 100041. DOI: 10.1016/j.caeai.2021.100041.
- [19] Tharapos M, Lau K H, Peszynski K, et al. Generative AI in accounting education: Evaluating ChatGPT's role in assessment and skill development. *Accounting & Finance*, 2025, 65(3): 3231-3244. DOI: 10.1111/acfi.70051.