

To cite this article: Gregorius Punto Aji and Angelina Kusuma Jelita Mawarni (2026). BEYOND THE “SAGE ON THE STAGE”: GENERATIVE AI AND THE CHALLENGING ROLES IN ESP TEACHING, International Journal of Education and Social Science Research (IJESSR) 9 (4): 273-288 Article No. 1270, Sub Id 1943

BEYOND THE “SAGE ON THE STAGE”: GENERATIVE AI AND THE CHALLENGING ROLES IN ESP TEACHING

Gregorius Punto Aji¹ and Angelina Kusuma Jelita Mawarni²

¹Universitas Sanata Dharma, Yogyakarta, Indonesia

²Sekolah Bunda Mulia, Jakarta, Indonesia

DOI : <https://doi.org/10.37500/IJESSR.2026.9411>

ABSTRACT

The massive spread of Generative Artificial Intelligence (Gen AI) into language classrooms has reopened a currently relevant question related with the human teacher's role, and in some quarters revived the worry that the instructor who once stood as the “Sage on the Stage” has become obviously less important. This narrative literature review asks what actually happens to the roles and professional identities of language teachers as Gen AI has massively intervened their work, with particular attention to English for Specific Purposes (ESP). Following a protocol-guided search of academic databases, thirteen peer-reviewed studies published between 2022 and 2026 were synthesized, read alongside a smaller body of abundant literature on general language teaching. Taken together, the studies point away from displacement and toward reinvention. Teachers, ESP practitioners in particular, are taking on the work of prompt design, drawing on what one study terms AI-pedagogical knowledge to translate tacit expertise into instructions a model can follow. Since Gen AI can produce credible but incorrect content in specific fields such as engineering, medicine, law, etc., ESP teachers also find themselves play roles as domain validators and as guides to critical AI literacy. What emerges is less an authoritative source of knowledge than a facilitator who decides when to lean on automation and when to rely on judgment, context, and rapport that a model cannot supply.

KEYWORDS: English for Specific Purposes (ESP), Generative AI, teacher identity, prompt engineering, critical AI literacy.

INTRODUCTION

Since abundant language learning sources became widely available, few developments have unsettled language teaching quite like Generative Artificial Intelligence (Gen AI), and the debate over what it means for the human teacher shows no sign of settling (Tang et al., 2024; Yan, 2023). It has become more complicated in English for Specific Purposes (ESP), in which teaching becomes more dependent on particular occupations, professions, and disciplines. If a model can produce a workable legal memo or a clinical case note on demand, is there anything left for the teacher who used to be the primary authority in the classroom - the “Sage on the Stage”?

Gen AI enables teachers to customize learning content in real time, provide formative feedback, and adjust assignments to suit various proficiency levels and learning contexts (Jeon & Lee, 2023; Moorhouse, 2024). In other words, Gen AI can help English teachers design innovative, engaging, and flexible learning experiences by overcoming the limitations of traditional pedagogical methods, such as a lack of personalized instruction, limited opportunities for collaboration, and inadequate automated material development (Ghafouri, 2024; Lee et al., 2025; Moorhouse, 2024).

ESP practitioners have never had a light workload. The job has long meant running careful need analyses and then building materials tailored to a discipline the teacher may not fully belong to (Dudley-Evans & St John, 1998; Flowerdew, 2023). Tools such as ChatGPT, Claude, and Gemini appear to shortcut much of this: They draft technical texts, stage professional dialogues, and sketch syllabi in seconds (Yan, 2023; Zebua, 2024). Reactions have varied. Some criticisms argue that teachers can become deskilled or edged out; others see that no model reproduces the situated judgment that a good teacher brings to a classroom.

Chen, Tawilapakul, and Lin (2026) offer a useful way through this impasse. Following one experienced ESP instructor over an extended period, they found that working with Gen AI did not replace her authority so much as expose a competence she had not previously had to name, which they call AI-pedagogical knowledge. Using a model well, it turned out, required her to unpack intuitive teaching decisions into explicit, ordered prompts. Generating text was the easy part; supplying the context and judgment behind it, what the authors term *teacherness*, was not something that the model could do on its own.

The field of ESP has evolved significantly in recent years, adapting to the demands of ever-changing global communication, technological advances, and the increasing specialization of various professional fields (Dubovyk, 2024). This is also where ESP parts company with general English. In EFL or general-purpose classes, Gen AI is mostly asked for everyday language work: grammar drills, vocabulary, generic writing, along with guided practice for interviews, presentations, and exams. Systematic reviews of classroom-based studies confirm this orientation, with Gen AI functioning largely as a resource provider for writing topics, grammar, and vocabulary work (Lee et al., 2025), and with commercial applications such as ELSA Speak targeting discrete, decontextualized skills like pronunciation accuracy (Abdelrady et al., 2026). These additional features further enrich the learning experience. ESP leans on specialized professional knowledge, and it is precisely in these high-stakes domains that documented error rates are most alarming. Dahl et al. (2024) found hallucination rates of 69–88% when leading models were asked specific, verifiable questions about federal court cases. Omar et al. (2025) embedded a single fabricated clinical detail in 300 physician-validated vignettes and found that six models elaborated on the false detail in 50–82% of outputs; a mitigation prompt lowered the mean rate from 66% to 44% but eliminated it in no model. Compounding this, models are poor judges of their own specialised output: Minder et al. (2025) report that ChatGPT's error-

annotation F_1 fell from 0.707 when evaluating another system's translations to 0.496 when evaluating its own, with false errors rising to 55%. That risk changes the teacher's job. The ESP practitioner is not only teaching language but checking output against professional reality, and teaching students to do the same before that output reaches a client, a patient, or a court. Jenks (2025) emphasizes the role of experts in intercultural communication in determining the appropriate use of Gen AI in language teaching, particularly by integrating the culturally sensitive communication strategies necessary for developing inclusive ELT teaching materials.

It has been admitted that technology has made the process of teaching and learning English for Specific Purposes (ESP) more engaging and enjoyable. However, the integration of Gen AI into English language teaching (ELT) also presents significant pedagogical challenges. Overreliance on Gen AI can discourage creativity and critical thinking, and may lead to unfairness in assessment, as students who use AI tools may gain an unfair advantage (Wiboolyasarin et al., 2024; Yan et al., 2024). Gen AI has the potential to transform language education by providing innovative approaches to language content creation, transcription, and pedagogical practices (Michaels et al., 2023). Gen AI models featuring such linguistic functions can increase the quality of access to learning materials using the native languages of students and their proficiency in foreign ones. Adere (2022) showed how AI can solve the problem of language inaccessibility to education, making it more diverse and popular.

For all the recent issues, it still remains relatively little discussion about how Gen AI reshapes the professional identity and day-to-day decisions of ESP teachers specifically; most reviews stay at the level of general English or student outcomes (Tang et al., 2024; Yan, 2023). Therefore, his review addresses that gap through three questions:

RQ1: What new roles and competencies are emerging for language teachers as Gen AI becomes integrated into instructional and material-development practices in ESP contexts?

RQ2: How does the domain-specific hallucination risk of Gen AI reposition ESP practitioners as critical validators and AI-literacy facilitators in specialized professional fields?

RQ3: In what ways is the professional identity of language teachers being restructured in response to Gen AI integration, and what psychological tensions characterize this transition?

2. METHOD

To answer the identified RQs, this study adopts a narrative literature review, an approach suited to an emerging and fast-moving field in which the evidence is still too uneven in design and focus for the statistical aggregation a systematic review assumes. Rather than pooling comparable outcomes, a narrative synthesis reads studies of different kinds against one another to trace how a phenomenon is being understood across a body of work, here, how the roles and professional identity of ESP teachers are being reshaped as generative AI enters their practice.

2.1 Searching Strategy and Sources

The search used the terms including “ESP”, “English for Specific Purposes,” “ESP teacher,” “Generative AI,” “ChatGPT,” “pedagogical shift,” and “teacher identity,” searching in Google Scholar, DOAJ, and ScienceDirect. Because the field moves very fast, the search was limited to articles published between 2022, when generative chatbots became widely available, and 2026. AI-assisted discovery tools were used to help surface candidate sources; every item they returned was then identified, read, and verified against its original publication before they were selected for inclusion.

2.2 Inclusion and Exclusion Criteria

Selected articles were included if they (a) were peer-reviewed journal articles or reputable conference papers, (b) were written in English and published between 2022 and 2026, (c) addressed generative AI in language teaching or learning, and (d) bore, directly or by clear extension, on the roles, competencies, or professional identity of teachers, with priority given to English for Specific Purposes, English for Professional Purposes, English for Occupational Purposes, English for Vocational Purposes, and English for Academic Purposes contexts. Meanwhile, others were excluded if they addressed learner outcomes with no bearing on the teacher’s role or identity, produced AI-generated sources outside language teaching, or had not undergone peer review, such as blog posts, opinion pieces, and unreviewed preprints.

2.3 Study Selection and Corpus

Titles and abstracts were selected on these criteria, and the remaining full texts were read in full and analyzed before a final decision was made. Thirteen studies met all criteria and formed the core corpus for synthesis, summarized in Table 1 (Appendix). Because research dealing specifically with ESP practitioners and generative AI remains thin, these were read alongside a smaller body of adjacent literature from EFL and general language-teacher education wherever it spoke directly to ESP concerns. The corpus spans a range of settings and designs, ethnography, qualitative case study, survey and other quantitative designs, mixed-methods intervention, systematic review, and conceptual or technology-review pieces.

2.4 Analytical Procedure

Analysis followed a thematic synthesis. The main findings of each study were first extracted and summarized; these were then coded and compared across the corpus to surface recurring patterns; and the patterns were consolidated into three themes that organize the discussion that follows: (1) the move from content provider to prompt designer, (2) the turn toward critical evaluation and AI-literacy facilitation, and (3) the reworking of teachers’ professional identity. Each study was also mapped to the research question or questions it primarily informed, as set out in Table 1 (Appendix).

3. FINDINGS

Across the studies on the thirteen selected articles, three themes stood out, and they are worth reading as a sequence: as the teacher's routine tasks shift, so does what counts as expertise, and eventually so does how teachers see themselves. The arguments to view AI-assisted teaching as a partner rather than a replacement is not new. Ji, Han, and Ko's (2022) systematic review of conversational AI came to a similar conclusion, before generative models emerged, that these efforts worked best in collaboration with human teachers, instead of replacing them.

3.1 From Content Provider to Prompt Engineer

The most obvious change that the literature records is in how ESP teachers develop their materials. Sourcing authentic discipline-specific texts, verifying them, and adapting them to a particular group of learners used to absorb a large share of preparation time (Flowerdew, 2023; Ngo & Hastie, 2025). Gen AI compresses much of that: a serviceable draft or a first-pass syllabus now takes seconds. What this does not work is to make the teacher redundant. Chen, Tawilapakul, and Lin (2026) describe that it requires a different competence, the *AI-pedagogical knowledge*. This competence has a recognizable lineage: it reads as a generative-AI inflection of the technological pedagogical content knowledge (TPACK) that has long described how teachers weave new tools into subject teaching (Mishra & Koehler, 2006), and scholars working directly on that framework now argue that generative models strain it, asking teachers for judgments about reliability and ethics that a static, teacher-controlled view of technology never had to make (Mishra, Warr, & Islam, 2023).

In practice this means teaching through the AI-assisted instructions. To get the linguistically accurate and appropriate materials for a specialized field, the ESP teachers have to make explicit the pedagogical reasoning that usually stays tacit, breaking expert judgment into a sequence the model can act on. Ngo and Hastie (2025) argue that this kind of AI literacy is no longer a nice-to-have; designing learning tasks, working with new teaching models, and guiding students all now assume it. The final result is that a model's output is only as good as the teacher steering it, which returns the human back to the center of the design process rather than to one side of it (Kohnke, Moorhouse, & Zou, 2023; Assassi & Rouaghe, 2025).

3.2 The Shift Toward Critical Evaluation and AI Literacy Facilitation

Once a model can produce clean technical implementations, such as business correspondence, reports, documentation, and many others, the point of ESP teaching moves up a level, from focusing on the grammar right to determining whether the content holds. Several of the reviewed articles have showed the same limitation: in specialized domains, models generate confident errors, whether invented terminology or claims that are simply inaccurate for the field (Tang et al., 2024; Yan, 2023; Assassi & Rouaghe, 2025). The productive response is not to ban the tools but to treat what they produce as a rough draft that has to be checked. Ngo and Hastie's (2025) Artificial Intelligence for Academic Purposes (AIAP) teaching models is an effort to give students a way of working along the line between using a model and passing its work off as their own.

This redefines the teacher's role, but this does not replace it - This redistributes it. Ji, Han, and Ko's (2022) systematic study is instructive here - mapping traditional language-teaching functions against what conversational AI could actually work. They found that models absorbed some of the more mechanical work, such as routine practice and certain kinds of feedback. Meanwhile, the roles that call for judgment suggest the teacher's role as facilitator, monitor, discussion leader, and evaluator. Generative AI models raise the ceiling on what the technology can produce, but they leave that division largely intact and arguably sharpen it, since output that looks finished still needs someone to decide whether it is.

The job, then, is less about offering grammar corrections than about enhancing students' ability to critically test and question an AI's output (Samsudin & Allen, 2026; Lim & Darwin, 2026), and, in ESP, to weigh it against how the discipline actually operates. Authenticity and authorship get renegotiated in the learning process - students are expected to comprehend, evaluate and defend text rather than literally receive it. To do this a teacher's guidance is inevitably vary important. That is what keeps them from outsourcing their thinking, and it is the answer to the RQ2. The tendency of Gen AI to fail in exactly the high-stakes fields ESP serves—medicine, law, engineering—is what makes the teacher's role as validator and literacy guide hard to do without.

3.3 Psychological Restructuring of Professional Teacher Identity

The third theme is harder to see but probably the most consequential: What all of this does to how teachers understand themselves. To be the classroom authority as the sole source of knowledge has long been part of the teacher's main role, and Gen AI redefine that directly. The studies report a familiar first reaction, insecurity of being deskilled, losing autonomy, running into institutional obstacles, or being made obsolete by a more capable machine (Zebua, 2024; Assassi & Rouaghe, 2025). Uncertainty about how AI will actually be implemented, and about where teachers fit once it is, tends to sharpen that unease early on (UNESCO, 2023). ESP teaching contexts show it plainly: the language-specialist teachers in Assassi and Rouaghe's (2025) study describes something close to a loss of confidence, self-doubt, situational anxiety, even embarrassment. This becomes most acute when a subject-matter gap opens up that the AI cannot reliably close either.

Setting against this, the recent work is reassuring on one point: building solid Gen AI literacy is what keeps automation from eliminating the human role of teaching (Park, 2025; Huang, Wu, & Wu, 2025). There is some urgency here, because over-reliance cuts both ways. Tian and Zhang (2025) relate the heavy learner dependence on AI with cognitive fatigue and weaker critical thinking, which is precisely the failure a mediating teacher exists to prevent. The identity that seems to emerge on the far side of the initial anxiety is that of a mediator: handing routine and administrative work to the model (Moorhouse & Wong, 2025) while holding on to the parts of teaching that are stubbornly human. Moving past the "sage on the stage" does not mean stepping back so much as redirecting attention to emotional support and local, contextual judgment that automated systems do not have.

Meanwhile, that transition is not available to everyone on equal terms. Davin et al.'s (2026) cross-national survey found that access to structured Gen AI professional development varies a great deal

by country and by school levels, with secondary and university teachers reporting substantially more training and use than their elementary and middle-school colleagues. In other words, the shift toward a mediator identity is itself unevenly distributed, shaped by who gets trained and who does not. This is the answer to RQ3 with a caution rather than a clean resolution.

4. DISCUSSION

Discussed separately, the three themes describe three changes, while read together, they describe one. The shift from content provider to prompt designer suggests the question of what expertise now consists of: Once a model can generate fluent domain text, the teacher's value moves from producing the language to judging it; and it is that migration, repeated across tasks, which first unsettles and then reshapes how teachers understand themselves. The common thread is a single repositioning: from the authority who supplies knowledge to the mediator who decides when to lean on the model and when to withhold it. What the literature documents, in other words, is not three parallel effects of generative AI but one shift observed at three depths: task, competence, and identity. Figure 1 sets out that repositioning across the three levels.

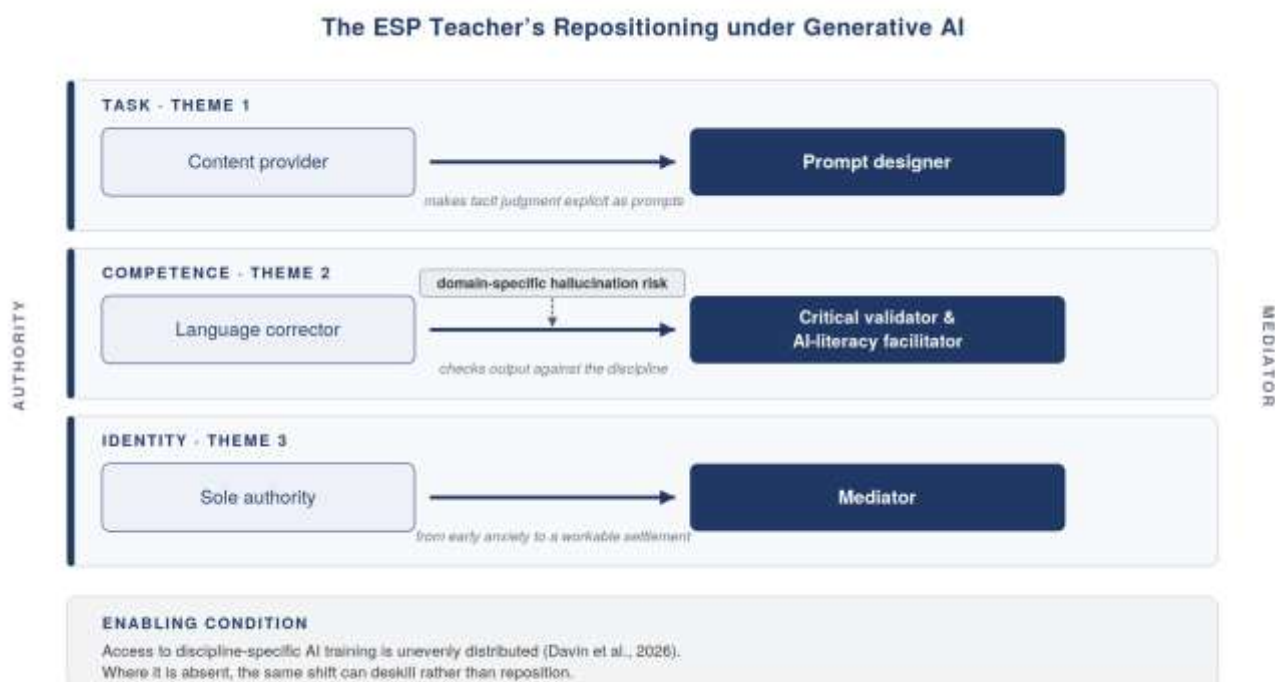


Figure 1: The ESP teacher's repositioning under generative AI. As routine tasks shift, so does what count as expertise, and in time so does professional identity; yet across all three levels the movement is the same, from an authority who dispenses knowledge to a mediator who decides when to lean on the model and when to rely on judgment it cannot supply. The tendency of generative models to fail in high-stakes fields is what turns the efficiency of prompting into an imperative to validate, and the shift holds only where teachers can reach the training that builds this competence.

Seen this way, the three research questions answer one another. RQ1 asks what new competencies emerge, and the corpus converges on AI-pedagogical knowledge (Chen, Tawilapakul, & Lin, 2026; Ngo & Hastie, 2025), the capacity to externalize tacit pedagogical judgment as instructions a model can act on. But that competence is not self-justifying; it matters because of RQ2. Precisely because generative AI models fail most confidently in the high-stakes domains ESP serves (Tang et al., 2024; Assassi & Rouaghe, 2025), the same teacher who designs the prompt must also validate the output, and teach students to do likewise (Ngo & Hastie, 2025; Lim & Darwin, 2026). Thus, prompt design and critical validation are two faces of one role rather than two separate roles. RQ3 then registers what it may result to play that role: The anxiety, self-doubt, and situational embarrassment that the studies have found (Assassi & Rouaghe, 2025; Zebua, 2024) are not incidental but the felt experience of a professional identity being rebuilt in real time. With Gen AI literacy the resource lets teachers reduce the anxiety rather than be governed by it (Park, 2025; Huang, Wu, & Wu, 2025).

Two cautions have appeared from this discussion. First, the reinvention narrative is conditional rather than automatic. The mediator identity presupposes ways to the training that develops AI-pedagogical knowledge, and Davin et al. (2026) show that access is distributed unequally by countries and school levels; where it is absent, the same technology may deskil the teacher rather than reposition them. The optimistic views and the deskilling worry are therefore not rival predictions so much as descriptions of two ends of an equity gradient. Second, the shift carries a risk the studies rarely name: If a teacher's expertise comes to live in prompting rather than in independent command of the domain, the very subject-knowledge gap that Assassi and Rouaghe (2025) describe could widen under the appearance of being solved, with the model masking rather than closing what the teacher does not know. Learner over-reliance already has a documented cost (Tian & Zhang, 2025); a teacher-side analogue is at least plausible and warrants attention.

These findings extend the earlier accounts rather than overturning them. The position that classroom AI is a partner rather than a replacement predates generative models, Ji, Han, and Ko (2022) reached it for conversational AI, but the reviewed work suggests generative capability sharpens the division of labor rather than blurring it: output that looks finished raises, rather than lowers, the premium on someone able to decide whether it is. That said the evidence remains early. Much of the corpus is small-scale, cross-sectional, or drawn from single cases and particular regions, and it describes an adoption moment that is still in motion. The repositioning traced here is thus best read as a well-supported direction of travel rather than a settled destination, a qualification the implications and recommendations below take up.

5. CONCLUSION

Across the thirteen studies, the evidence points not to the teacher's obsolescence but to the teacher's remaking. Generative AI does not remove the human from ESP so much as reposition her, from the authority who supplies knowledge to the mediator who pairs what automation does well with the judgment, context, and rapport it cannot supply. That figure, at once a designer of prompts, a validator of output, and a guide to critical AI literacy, is harder to automate than the sage on the stage ever was,

precisely because generative models fail most visibly in the high-stakes fields ESP serves. The picture that emerges is cautiously optimistic rather than triumphant: a role being rebuilt, and rebuilt unevenly, but not retired.

5.1 Pedagogical Implications

Several practical points follow for teachers, teacher educators, and the people who set policy. Professional development needs to move past generic digital-skills training toward discipline-specific AI literacy: prompt design, critical reading of AI-generated domain content, and the ethics of using these tools in professional fields. ESP course designers can build explicit critical-AI-literacy frameworks into the curriculum, Ngo and Hastie's (2025) AIAP is one workable model, so that learners leave able to weigh AI output against real professional standards. Institutions should also take the psychological side seriously, offering mentoring and communities of practice for teachers working through the role uncertainty this period brings. Moreover, because training is not evenly distributed - Davin et al. (2026) show elementary and middle-school teachers consistently underserved next to secondary and university colleagues - support should be aimed deliberately at those groups rather than spread as if everyone started from the same place.

5.2 Limitations

A few caveats apply. The search covered three databases (Google Scholar, DOAJ, and ScienceDirect) and so left out major repositories such as Scopus, ERIC, and Web of Science, which might have surfaced further studies. Every source was verified independently, but the possibility of selection bias remains. The corpus also leans toward Asian and Western institutions, Davin et al. (2026), which drew on Colombian data, is the main exception, so the findings travel less well to underrepresented regions across Africa, the Middle East, and much of Latin America. Finally, the technology keeps moving, and studies from 2022 to 2024 may already describe early-adoption patterns that no longer match current practice.

5.3 Recommendations for Future Research

Two directions seem especially worth pursuing. Longitudinal work would help clarify how sustained collaboration with AI affects teachers' professional identity and how long they stay in the profession, questions a snapshot cannot answer. And finer-grained study is needed of how different branches of ESP - English for medical purposes versus English for legal purposes, say - place different demands on AI-pedagogical knowledge, since "the ESP teacher" is not a single figure.

REFERENCES

- Abdelrady, A. H., Akram, H., et al. Harnessing AI for pronunciation development: Investigating the impact of ELSA Speak on EFL learners' oral proficiency. *World Journal of English Language*, 16(4), 311. <https://doi.org/10.5430/wjel.v16n4p311>
- Adere, E. M. (2022) Blockchain in healthcare and IoT: A systematic literature review. *ScienceDirect Array*, 14, 100139, 1-20. <https://doi.org/10.1016/j.array.2022.100139>

- Assassi, T., & Rouaghe, F. (2025). Exploring subject knowledge gap in ESP teaching: Perspectives of language and subject specialist teachers on AI integration. *Teaching English Language*, 19(1), 185–225. <https://doi.org/10.22132/tel.2024.460057.1621>
- Chen, Q., Tawilapakul, U., & Lin, A. M. Y. (2026). Teacherness in the age of Gen AI: An ethnographic exploration of pedagogical decision-making and AI integration in English language teaching. *System*, 136, 103877. <https://doi.org/10.1016/j.system.2025.103877>
- Dahl, M., Magesh, V., Suzgun, M., & Ho, D. E. (2024). Large legal fictions: Profiling legal hallucinations in large language models. *Journal of Legal Analysis*, 16(1), 64–93. <https://doi.org/10.1093/jla/laae003>
- Davin, K. J., Kissau, S., Haudeck, H., Ade-Thurrow, B., Gómez-Pereira, D., Oon, P. T., & Dornburg, A. (2026). Global and educational disparities in AI integration: A study of L2 teacher training and usage patterns. *System*, 136, 103901. <https://doi.org/10.1016/j.system.2025.103901>
- Dubovyk, O. (2024). Innovations in ESP teaching and learning practices. *Humanities Studies: Pedagogy, Psychology, Philosophy*, 12(4), 23–33. <https://doi.org/10.31548/hspedagog/4.2024.23>
- Dudley-Evans, T., & St John, M. J. (1998). *Developments in English for specific purposes: A multi-disciplinary approach*. Cambridge University Press.
- Flowerdew, L. (2023). *Needs analysis and ESP syllabus design*. Routledge.
- Ghafouri, M. (2024). ChatGPT: The catalyst for teacher-student rapport and grit development in L2 class. *System*, 120, Article 103209. <https://doi.org/10.1016/j.system.2023.103209>
- Huang, T., Wu, C., & Wu, M. (2025). Developing pre-service language teachers' Gen AI literacy: An interventional study in an English language teacher education course. *Discover Artificial Intelligence*, 5, Article 53. <https://doi.org/10.1007/s44163-025-00435-1>
- Ji, H., Han, I., & Ko, Y. (2022). A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. *Journal of Research on Technology in Education*, 1–19. <https://doi.org/10.1080/15391523.2022.2142873>
- Jenks, C. J. (2025). Communicating the cultural other: Trust and bias in generative AI and large language models. *Applied Linguistics Review*, 16(2), 787–795. <https://doi.org/10.1515/applirev-2024-0196>
- Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, 28(12), 15873–15892. <https://doi.org/10.1007/s10639-023-11834-1>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550. <https://doi.org/10.1177/00336882231162868>

- Lee, S., Choe, H., Zou, D., & Jeon, J. (2025). Generative AI (Gen AI) in the language classroom: A systematic review. *Interactive Learning Environments*, 335–359. <https://doi.org/10.1080/10494820.2025.2498537>
- Lee, S., Jeon, J., & Choe, H. (2025). Generative AI (Gen AI) and pre-service teacher agency in ELT. *ELT Journal*, 79(2), 287–296. <https://doi.org/10.1093/elt/ccaf005>
- Lim, K. Y., & Darwin, R. (2026). Critical digital literacies, generative AI, and the negotiation of agency in human-AI interactions. *System*, 136, 103904. <https://doi.org/10.1016/j.system.2025.103904>
- Michaels, H., Rinderle, M., Benesperi, I., Freitag, R., Gagliardi, A., & Freitag, M. (2023). Emerging indoor photovoltaics for self-powered and self-aware IoT towards sustainable energy management. *Chemical Science*, 14(20), 5350-5360. <https://doi.org/10.1039/d3sc00659j>
- Minder, J., Wisniewski, G., & Kübler, N. (2025). Testing LLMs' capabilities in annotating translations based on an error typology designed for LSP translation: First experiments with ChatGPT. In *Proceedings of Machine Translation Summit XX: Volume 1* (pp. 190–203). <https://aclanthology.org/2025.mtsummit-1.16/>
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and generative AI. *Journal of Digital Learning in Teacher Education*, 39(4), 235–251. <https://doi.org/10.1080/21532974.2023.2247480>
- Moorhouse, B. L. (2024). Generative artificial intelligence and ELT. *ELT Journal*, 78(4), 378–392. <https://doi.org/10.1093/elt/ccae032>
- Moorhouse, B. L., & Wong, K. M. (2025). *Generative artificial intelligence and language teaching* (Elements in Generative AI in Education). Cambridge University Press. <https://doi.org/10.1017/9781009618823>
- Ngo, T. N., & Hastie, D. (2025). Artificial Intelligence for Academic Purposes (AIAP): Integrating AI literacy into an EAP module. *English for Specific Purposes*, 77, 20–38. <https://doi.org/10.1016/j.esp.2024.09.001>
- Omar, M., Sorin, V., Collins, J. D., Reich, D., Freeman, R., Gavin, N., Charney, A., Stump, L., Bragazzi, N. L., Nadkarni, G. N., & Klang, E. (2025). Multi-model assurance analysis showing large language models are highly vulnerable to adversarial hallucination attacks during clinical decision support. *Communications Medicine*, 5(1), 330. <https://doi.org/10.1038/s43856-025-01021-3>

- Park, J. (2025). A systematic literature review of generative artificial intelligence (Gen AI) literacy in schools. *Computers and Education: Artificial Intelligence*, 9, 100487. <https://doi.org/10.1016/j.caeai.2025.100487>
- Samsudin, & Allen, C. J. (2026). Integrating generative AI in English education: Opportunities and challenges for EFL learners. *Journal of Research in English Language Teaching and Linguistics (JRELL)*, 2(1), 49–66. <https://doi.org/10.65431/jrell.v2i1.32>
- Tang, K.-S., Cooper, G., Rappa, N., Cooper, M., Sims, C., & Nonis, K. (2024). A dialogic approach to transform teaching, learning and assessment with generative AI in secondary education: A proof of concept. *Pedagogies: An International Journal*, 19(3), 493–503. <https://doi.org/10.13140/RG.2.2.20564.41608>
- Tian, J., & Zhang, R. (2025). Learners' AI dependence and critical thinking: The psychological mechanism of fatigue and the social buffering role of AI literacy. *Acta Psychologica*, 260, 105725. <https://doi.org/10.1016/j.actpsy.2025.105725>
- UNESCO. (2023). *Guidance for generative AI in education and research*. United Nations Educational, Scientific and Cultural Organization. <https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>
- Wiboolyasarini, W., Wiboolyasarini, K., Suwanwihok, K., Jinowat, N., & Muenjanchoey, R. (2024). Synergizing collaborative writing and AI feedback: An investigation into enhancing L2 writing proficiency in wiki-based environments. *Computers and Education: Artificial Intelligence*, 6, 100228. <https://doi.org/10.1016/j.caeai.2024.100228>
- Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, 28, 13943–13967. <https://doi.org/10.1007/s10639-023-11742-4>
- Zebua, N. (2024). Evaluating the impact of Gen AI in high school education (A critical review). *Polygon: Jurnal Ilmu Komputer dan Ilmu Pengetahuan Alam*, 2(6), 79–86. <https://doi.org/10.62383/polygon.v2i6.296>

APPENDIX

Table 1. Summary of Selected Studies (n = 13)

No.	Study (Author, Year)	Context & setting	Design / Data	Key Findings & Relevance to RQs	Theme(s)
1	Assassi & Rouaghe (2025)	ESP; perceptions of language-specialist vs. subject-specialist (non-native) teachers on AI integration	Mixed-methods (surveys + interviews)	Documents the subject-knowledge gap; language-specialist teachers voice a crisis of confidence (self-doubt, situational anxiety, embarrassment), sharpest when a subject-matter gap opens that AI cannot reliably close; a cautious stance toward ChatGPT for bridging domain gaps. → RQ2, RQ3	2, 3
2	Chen, Tawilapakul & Lin (2026)	ELT/ESP; one experienced ESP practitioner observed over an extended period	Ethnography (single case)	Gen AI does not erode teacher authority but surfaces a new competence — AI-pedagogical knowledge; the teacher translates tacit expertise into explicit prompts; “teacherness” (context + judgment) cannot be replicated by the model. → RQ1, RQ3	1, 3
3	Davin et al. (2026)	Cross-national (incl. Colombian data); L2 teachers across school levels	Cross-national survey	Access to structured Gen AI professional development varies widely by country and school level; primary/middle-school teachers are consistently underserved relative to secondary and university peers. → RQ3 (equity caveat)	3
4	Huang, Wu & Wu (2025)	Pre-service English language teacher education	Interventional study	Building robust Gen AI literacy is what keeps automation from eroding the human core of teaching. → RQ3	3

No.	Study (Author, Year)	Context & setting	Design / Data	Key Findings & Relevance to RQs	Theme(s)
5	Ji, Han & Ko (2022)	Language education; conversational AI (pre-generative)	Systematic review	Conversational AI works best as a collaborator, not a replacement; mechanical roles (routine practice, some feedback) are absorbed by AI, while judgment-based roles (facilitator, monitor, discussion leader, evaluator) remain with the teacher. → RQ1, RQ2	1, 2 (framing)
6	Kohnke, Moorhouse & Zou (2023)	Language teaching & learning (general)	Technology review (conceptual)	Output quality depends on the teacher steering the model; places the human at the center of the design process rather than to one side of it. → RQ1	1
7	Lim & Darvin (2026)	Multilingual secondary-school students, Canada; critical digital literacies; human–AI interaction	Qualitative case study (interviews, observations, digital artifacts; thematic & multimodal analysis)	Emphasizes critical digital literacies and the negotiation of agency; learners are positioned to evaluate and defend texts rather than accept them uncritically. → RQ2	2
8	Ngo & Hastie (2025)	EAP; 10-week AI- integrated module at a pathway college, Scotland	Framework development (AIAP) + mixed- methods intervention (pre/post)	The AIAP framework integrates AI literacy into the module; AI literacy becomes a prerequisite for task design and guidance, giving students a way to work along the line between using a model and passing off its work as their own; improved critical evaluation and ethical use. → RQ1, RQ2	1, 2

No.	Study (Author, Year)	Context & setting	Design / Data	Key Findings & Relevance to RQs	Theme(s)
9	Park (2025)	Schools; Gen AI literacy	Systematic literature review	Building Gen AI literacy is what protects the human core of teaching from erosion by automation. → RQ2, RQ3	2, 3
10	Samsudin & Allen (2026)	Undergraduate EFL learners; everyday English learning with Gen AI	Interpretive qualitative design; 20 undergraduate EFL learners; semi-structured interviews + reflective journals (4 weeks); thematic analysis	Learners perceive Gen AI as a double-edged tool: it supports confidence, flexibility, and engagement through instant feedback, reduced anxiety, and help with vocabulary and ideas, but raises overreliance, weakened critical thinking, surface-level understanding, and concerns over inaccurate output and ethics; the study urges guided, reflective use and casts the educator as the one who builds critical awareness. → RQ2, RQ3	2, 3
11	Tian & Zhang (2025)	University students (China); psychological mechanisms of AI use	Quantitative survey (n = 580; moderated-mediation model; cross-sectional)	Greater AI dependence is associated with weaker critical thinking, partly mediated by cognitive fatigue; AI/information literacy buffers the effect — the very failure a mediating teacher exists to prevent. → RQ2, RQ3	3 (and 2)
12	Yan (2023)	L2 writing practicum	Exploratory investigation	Shows Gen AI can shortcut writing work but also produces confident errors that must be checked; underscores the need for human judgment over model output. → RQ1, RQ2	1, 2

No.	Study (Author, Year)	Context & setting	Design / Data	Key Findings & Relevance to RQs	Theme(s)
13	Zebua (2024)	High-school education; critical review of Gen AI	Critical review	Highlights deskilling and obsolescence concerns and the risks of Gen AI; records teachers' early anxiety toward the technology. → RQ3	3

Note. RQ1 = new roles and competencies (content provider → prompt designer); RQ2 = teacher as critical validator and AI-literacy facilitator; RQ3 = restructuring of professional identity.