





RESEARCH ARTICLE

# Generative AI and English language teaching: A global Englishes perspective

Seongyong Lee<sup>1</sup> , Jaeho Jeon<sup>2</sup> , Jim McKinley<sup>3</sup>  and Heath Rose<sup>4</sup> 

<sup>1</sup>School of Education and English, University of Nottingham Ningbo China, Ningbo, China; <sup>2</sup>Department of Curriculum and Instruction, The University of Alabama, Tuscaloosa, AL, USA; <sup>3</sup>IOE – Culture, Communication & Media, University College London, London, UK and <sup>4</sup>Department of Education, Oxford University, Oxford, UK

**Corresponding author:** Jaeho Jeon; Email: [jaehojeon21@gmail.com](mailto:jaehojeon21@gmail.com)

## Abstract

Generative AI (GenAI) offers potential for English language teaching (ELT), but it has pedagogical limitations in multilingual contexts, often generating standard English forms rather than reflecting the pluralistic usage that represents diverse sociolinguistic realities. In response to mixed results in existing research, this study examines how ChatGPT, a text-based generative AI tool powered by a large language model (LLM), is used in ELT from a Global Englishes (GE) perspective. Using the Design and Development Research approach, we tested three ChatGPT models: Basic (single-step prompts); Refined 1 (multi-step prompting); and Refined 2 (GE-oriented corpora with advanced prompt engineering). Thematic analysis showed that Refined Model 1 provided limited improvements over Basic Model, while Refined Model 2 demonstrated significant gains, offering additional affordances in GE-informed evaluation and ELF communication, despite some limitations (e.g., defaulting to NES norms and lacking tailored GE feedback). The findings highlight the importance of using authentic data to enhance the contextual relevance of GenAI outputs for GE language teaching (GELT). Pedagogical implications include GenAI–teacher collaboration, teacher professional development, and educators’ agentic role in orchestrating diverse resources alongside GenAI.

**Keywords:** generative AI; GenAI; English language teaching; global Englishes; teacher agency

GenAI is reshaping ELT by enhancing both teaching methodologies and student engagement, but it requires careful integration into the language classroom. GenAI is an umbrella term encompassing pre-trained AI technologies, including image, music, and text generation tools. LLMs serve as the foundation for text-based AI tools

such as ChatGPT, Claude, and Gemini, which process big linguistic data to generate human-like text (Moorhouse, 2024). Recent studies highlight that GenAI tools can promote student autonomy (Van Horn, 2024), personalized learning (Xiao & Zhi, 2023), and collaborative learning (Ghafouri, 2024). Specifically, GenAI serves diverse roles in ELT, such as content curator (Karataş et al., 2024; Lee et al., 2023), evaluative feedback provider (Ghafouri, 2024), and conversation partner (Jeon & Lee, 2023; Van Horn, 2024). By supporting instructional design and material development through the automation of content and task generation tailored to learners' needs (Choi et al., 2024; Lee et al., 2023), LLM-based GenAI allows teachers to customize instructional content in real time, provide formative feedback, and differentiate tasks to accommodate diverse proficiency levels and learning contexts (Jeon & Lee, 2023; Moorhouse, 2024). Therefore, GenAI can help ELT teachers design innovative, engaging, and adaptable learning experiences by addressing limitations in traditional pedagogical methods, such as a lack of personalized instruction, limited collaborative opportunities, and insufficient automated material development (Ghafouri, 2024; Lee et al., 2025b; Moorhouse, 2024). However, the integration of GenAI into ELT also presents significant pedagogical challenges. Overreliance on GenAI may hinder creativity and critical thinking and may lead to assessment inequities as students using AI tools could gain an unfair advantage (Wiboolyasarin et al., 2024; Yan et al., 2024). Also, concerns about the quality and originality of GenAI-generated content, which can often lack depth and specificity, pose further pedagogical obstacles (Abdelhalim, 2024; Jeon & Lee, 2023).

LLM-based GenAI tools may reinforce ELT hierarchies. Biases in training data that overrepresent dominant linguistic norms while marginalizing minority voices (Jeon et al., 2025; Navigli et al., 2023; Payne et al., 2024) can reinforce stereotypes, limit intercultural understanding, and undermine the lingua-cultural diversity essential for developing critical linguistic competence (Lee et al., 2025a; Payne et al., 2024). These biases arise from data representation imbalances and algorithmic standardization (Bender & Koller, 2020; Hovy & Prabhumoye, 2021). The overrepresentation of dominant languages leads to outputs that reflect mainstream norms, thereby marginalizing diverse linguistic practices (Dai et al., 2024; Payne et al., 2024). Furthermore, LLMs' statistical algorithms often result in homogenization simplifying linguistic diversity and excluding less-resourced languages and their associated cultures (Bender et al., 2021; Dai et al., 2024). As a result, biased outputs can perpetuate monolingual ideologies, reinforce societal inequities, and undermine linguistic justice in L2 education (Dai et al., 2024; Jeon et al., 2025; Payne et al., 2024). Addressing these challenges requires the development of more diverse training datasets and better algorithms, despite ongoing resource limitations (Brandt & Hazel, 2025; Zhu et al., 2025). For example, Dai et al. (2024) highlight that integrating datasets with diverse accents, dialects, and interactional norms can help reduce the overrepresentation of dominant language varieties and mitigate sociolinguistic bias in AI-generated content. Likewise, Jenks (2025) emphasizes the role of intercultural communication scholars in defining the appropriate GenAI usage in language teaching, particularly by incorporating culturally sensitive communication strategies necessary for developing inclusive ELT materials.

Research on diversifying training datasets and refining algorithmic designs remains in its early stages (Brandt & Hazel, 2025; Dai et al., 2024), and their practical application

in multilingual ELT classrooms is still limited (Lee et al., 2025a; Lo, 2025). The use of LLMs to represent diverse English forms, especially from a pluricentric perspective, has largely been explored conceptually or technically rather than pedagogically (e.g., Dai et al., 2024). Thus, studies examining the pedagogical usefulness of LLM-powered GenAI for ELT from a sociolinguistic standpoint remain scarce. Moreover, prompt engineering has not been systematically investigated within a pluricentric ELT framework, underscoring the need for empirical research on adapting GenAI for practical classroom use. This article addresses these gaps by exploring how ChatGPT can support teachers in designing practical, GE-informed ELT materials applicable to classroom contexts. Specifically, it examines how two major sources of LLM bias – overrepresentation of native English varieties in training datasets and algorithmic standardization favoring dominant norms – can be mitigated through accessible prompting techniques. Rather than aiming to eliminate these biases at the technological level, this study focuses on the pedagogical usability of LLM tools in ELT by assessing teacher-friendly prompt engineering methods to adapt ChatGPT’s outputs to GELT-based classroom applications (Choi et al., 2024; Lee et al., 2024a).

Given the multilingual and intercultural nature of English, GE provides a paradigm that captures its linguistic, sociolinguistic, and sociocultural diversity (Rose & Galloway, 2019). GE is defined as “a critical paradigm that explores the evolving use of English as a global language by a diverse community of users” (Rose & McKinley, 2025, p. 3). This study adopts the GELT framework to examine how customized ChatGPT can support ELT task and material design.

The GELT framework derives from six proposals for ELT reform situated in World Englishes (WE – localized English varieties), English as a lingua franca (ELF – communication in English among speakers of different linguistic and cultural backgrounds), and English as an international language (EIL – English as a global medium for intercultural communication). It promotes diverse perspectives on English use, arguing that an exclusive focus on standard forms inadequately prepares students for global communication (Galloway & Rose, 2015; Rose & Galloway, 2019). Within this pluricentric framework, four core GELT principles relevant to GenAI in ELT include (1) exposing learners to diverse English varieties, (2) raising GE awareness, (3) respecting diverse cultures and identities, and (4) fostering awareness of ELF strategies while moving away from native-English speaker (NES) norms (Rose & Galloway, 2019). This shift challenges entrenched native-speaker norms, advocating for ELT to embrace local, context-specific knowledge and pluricentric English varieties (Kumaravadivelu, 2012).

This study examines how customized ChatGPT models can support GELT in three key roles: (1) a content curator generating inclusive teaching materials (Lee, 2020), (2) an evaluative feedback provider promoting diverse linguistic outcomes (Hu, 2012; Tsagari et al., 2023), and (3) a conversation partner facilitating ELF interactions (Boonsuk et al., 2024). ChatGPT customization involves the use of prompt engineering and the integration of GE-informed external data to mitigate biases inherent in its training data. By designing targeted prompts and incorporating context-specific datasets (e.g., GE corpus data), ChatGPT outputs can better reflect diverse English varieties and linguistic features.

This study proposes a three-stage framework that employs prompt engineering and GE-informed data sources (Sahoo et al., 2024; Vatsal & Dubey, 2024). The theoretical

grounding is drawn from instructional design principles (Choi *et al.*, 2024) and the Design and Development Research (DDR) approach (Richey & Klein, 2005, 2007).

## A case study: Developing GenAI-GELT instructional module

### *Design and development research methodology*

The DDR approach (Richey & Klein, 2005, 2007) was used to design and evaluate customized GenAI models, specifically an LLM-based GenAI tool such as ChatGPT, that support GELT, namely “GenAI-GELT instructional module” (see GenAI-GELT, 2024, for full prompts and responses). This study uses ChatGPT as an example of an LLM tool to examine its capacity for instructional design, while maintaining GenAI as an inclusive term to highlight its broader potential for ELT. Given that the DDR method is useful for crafting and examining instructional designs, tools, and programs (Lee *et al.*, 2024a), future research can adapt similar methodologies to evaluate diverse GenAI models, including but not limited to LLMs, in educational contexts. The term “module” refers to an instructional design unit that integrates a GenAI tool, data sources, and prompt engineering techniques to create, customize, and evaluate ELT tasks and materials – a set of prompts and GPT outputs grounded in ChatGPT models (Choi *et al.*, 2024; Lee *et al.*, 2024a). Building on instructional development research, DDR serves as a research methodology that designs and assesses “new procedures, techniques, and tools” through a systematic analysis of a specific case (Richey & Klein, 2005, p. 24). In this study, DDR provides a structured framework for iteratively refining and assessing instructional materials to ensure their alignment with GELT principles. As an evaluation tool, DDR allows the assessment of GenAI-generated ELT materials by analyzing if its outputs fit the practical classroom use from the GELT perspective (Richey & Klein, 2005; Rose & Galloway, 2019). Thus, DDR was expected to provide a systematic analysis of designing and discussing the quality of ChatGPT-based automatic instructional system for GELT (Lee *et al.*, 2024a).

### *Tool and instruments*

We utilized ChatGPT-4o, the most recent version at the time of data generation, for this study. ChatGPT-4o has improved capabilities with multilingual data, equipped with efficient multimodal functionalities, capable of handling text, audio, and image data, and outputting both text and speech (OpenAI, 2024). Also, ChatGPT-4o’s recent function of memory management allows for archiving conversation data across multiple sessions, thereby enabling users to obtain highly contextualized outputs.

To design and refine the GenAI-GELT instructional module, diverse prompt engineering techniques were adopted, such as single-step or multi-step prompting, prompt optimization, retrieval augmented generation (RAG), instructed prompting, and few-shot learning (Sahoo *et al.*, 2024; Vatsal & Dubey, 2024). These techniques were utilized in a complementary way to maximize its effectiveness in facilitating the capability of the module in generating quality responses (Vatsal & Dubey, 2024). Table 1 illustrates prompting engineering techniques used for developing the GenAI-GELT instructional module.

**Table 1.** Prompt engineering techniques used for the GenAI-GELT instructional module.

Techniques	Definition	Benefits and limitations
Single-step prompting	A single query method without additional sub-steps	Simple and efficient for tasks; limited in complex ones
Multi-step prompting	Sequential prompts with additional steps	Enhancing reasoning/accuracy; requiring complex design
Prompt optimization	Prompt adjustment for improved performance	Boosting efficiency; potentially inducing biases
Retrieval augmented generation (RAG)	Retrieving relevant information from an external knowledge source	Improving accuracy by using external knowledge; dependent on the database quality
Instructed prompting	Using explicit instructions by ignoring irrelevant context	Enhancing precision by screening distractions; potentially less flexible
Few-shot learning	Providing a few examples in addition to prompts	Effective in complex tasks; example selections may cause biases

**Table 2.** ChatGPT roles, GELT tasks and materials, and GELT proposals.

ChatGPT roles	GELT tasks and materials	GELT proposals
Task 1: Content curator	<ul style="list-style-type: none"> <li>■ Generating reading materials that reflect a specific variety of English</li> </ul>	<ul style="list-style-type: none"> <li>■ Exposure to varieties of English</li> </ul>
Task 2: Evaluative feedback provider	<ul style="list-style-type: none"> <li>■ Assessing and offering feedback on an L2 essay from Global Englishes</li> </ul>	<ul style="list-style-type: none"> <li>■ Raising GE awareness</li> <li>■ Respecting diverse cultures and identity</li> </ul>
Task 3 and 4: Conversation partner	<ul style="list-style-type: none"> <li>■ Text-based role-play with ChatGPT for ELF interaction</li> <li>■ Speech-based role-play with ChatGPT for ELF interaction</li> </ul>	<ul style="list-style-type: none"> <li>■ Raising GE awareness</li> <li>■ Raising awareness of ELF strategies</li> </ul>

We employed ChatGPT's three roles – content curator, evaluative feedback provider, and conversation partner – to assist teachers in designing GE-informed ELT tasks and materials: (1) generating WE materials (Galloway & Rose, 2018; Rose & Galloway, 2019; Rose et al., 2021); (2) providing evaluative feedback on ELF user essays (Hu, 2012; Tsagari et al., 2023); (3) facilitating ELF interaction (Boonsuk et al., 2024; Lee et al., 2025a). Rather than targeting students directly, these roles help teachers integrate GE principles into instructional design for inclusive, effective classrooms.

Table 2 shows the association among ChatGPT roles, GELT instructional tasks, and the four core proposals of GELT. The first task had ChatGPT curate content to raise awareness of Korean English, addressing the lack of GE-aligned materials (Rose et al., 2021). This variety was selected due to the research's local context, providing a practical example of GELT principles in localized English forms, with broader applicability to other varieties. The second task involved ChatGPT providing feedback on a Korean student's writing, testing whether prompt engineering could generate outputs sensitive to non-standard varieties. Instead of data fine-tuning, targeted prompts improved ChatGPT's responses (Lee et al., 2024a).

The third task guided ChatGPT as a conversation partner to help learners practice ELF strategies – repetition, clarification requests, and accommodation – key to mutual understanding in multilingual contexts (Seidlhofer, 2011). Rooted in GELT’s focus on communicative competence (Galloway & Rose, 2015), this task centered on Korean and Chinese English use. The research team, familiar with Asian Englishes, provided a practical context for evaluating ChatGPT’s support for ELF communication. Task Three involved ELF text-based interaction between an L1 Chinese and L1 Korean speaker, while Task Four examined spoken interaction. The choice of Korean and Chinese interlocutors reflected the first two authors’ national and research backgrounds, while the third author, a Global Englishes expert, had previously researched Korean English intelligibility to Chinese listeners.

To enhance the ChatGPT models, we incorporated three external data sources using the RAG technique (Table 1). My GPT configuration allowed the integration of custom datasets without modifying ChatGPT’s internal structure, enabling it to retrieve task-specific data during content generation (Lee et al., 2024b). In addition, instructed prompting techniques (Table 1) guided ChatGPT’s outputs to align with the linguistic and contextual features of the external datasets (Kim & Lu, 2024).

First, to optimize ChatGPT’s role as a content curator, we utilized the Gachon Learner Corpus (Carlstrom & Price, 2014). This corpus, comprising over 2.5 million tokens from 25,073 English essays by Korean students, provides authentic written representations of Korean English. Second, to establish GE-informed essay assessment rubrics for ChatGPT’s evaluative feedback role, we selected five publications on WE and ELF assessment (Cooke, 2020; Hu, 2012; Huda & Irham, 2023; Nakamura, 2020; Tsagari et al., 2023). These studies were identified through a SCOPUS search using keywords related to WE, ELF, and GE-informed assessment, focusing on explicit rubrics and assessment frameworks. Finally, to support ChatGPT as a conversation partner, we adopted the Corpus of English as a Lingua Franca in Academic Settings (ELFA corpus) (ELFA, 2008) to create ELF conversation scenarios and design a text-based chatbot. This one-million-word corpus, derived from spoken academic ELF discourse (seminars, lectures, and panels) was selected for its multidisciplinary focus in academic settings (Mauranen et al., 2010). With speakers from 51 L1 backgrounds, including significant contributions from Asian speakers, the largest proportion (28.5%), comes from Finnish English, influencing this study’s focus due to its structured ELF interactions.

### *Development and evaluation process*

The GenAI-GELT instructional module (GenAI-GELT, 2024) was developed iteratively using three ChatGPT models – Basic, Refined 1, and Refined 2 – following the DDR methodology (Lee et al., 2024a; Richey & Klein, 2005, 2007). The models were designed utilizing prompt engineering techniques, evaluated through a GELT perspective, and refined for GELT principles (Table 2). Systematic comparison of these models, as in previous DDR-based studies (Kim & Lu, 2024; Lee et al., 2024a, 2024b), helps assess prompt engineering’s effectiveness in mitigating linguistic biases and enhancing GELT-based ELT applications.

As illustrated in Figure 1, Basic Model used a single-step prompting method to generate four GELT instructional tasks and materials (Tables 1 and 2). It relied solely on

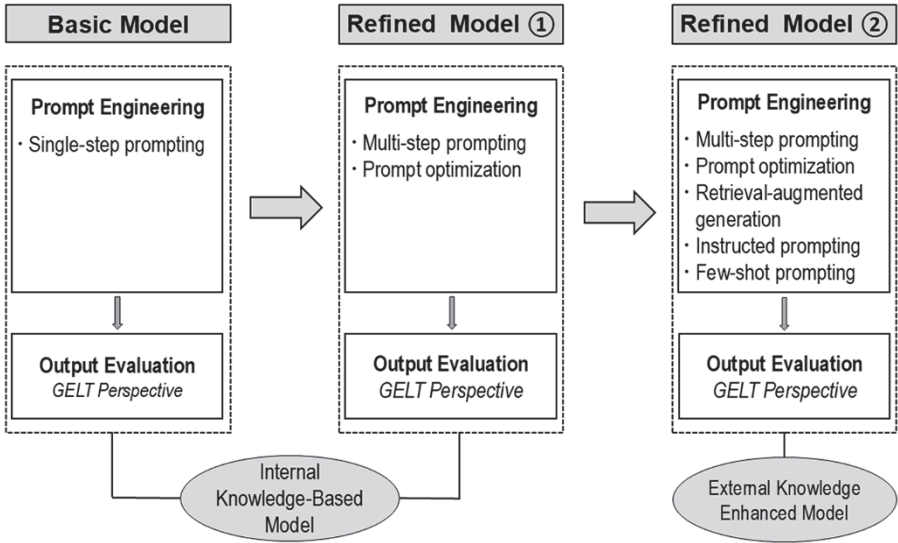


Figure 1. Research process of developing GenAI-GELT instructional modules.

ChatGPT's internal data, and outputs were analyzed against the four GELT proposals (Richey & Klein, 2005; Rose & Galloway, 2019; Rose & McKinley, 2025). Refined Model 1 adopted a multi-step prompting technique and prompt optimization, replacing the single-step approach. For example, in Task 2, instead of evaluating an essay without rubrics (as in the Basic Model), this version first generated writing rubrics based on ChatGPT's internal data and GELT principles, then applied them to essay assessment.

Refined Model 2 introduced three additional techniques, such as RAG, instructed prompting, and few-shot learning. Unlike the earlier models, which relied solely on ChatGPT's internal knowledge, this model integrated three external datasets:

- (1) the Gachon Learner Corpus for Korean English representation;
- (2) five GE-informed assessment publications for writing rubrics;
- (3) ELFA corpus for ELF conversation modeling.

These datasets were uploaded via the "knowledge" option in My GPT, allowing ChatGPT to retrieve task-specific data during content generation (Lee et al., 2024b). Instead of retraining ChatGPT-4o, the study optimized models using prompt-based techniques to create teacher-accessible solutions. For example, in Task 1, ChatGPT generated a sample essay reflecting Korean English using the Gachon corpus. In Task 2, it analyzed GE-informed assessment studies to generate and apply writing rubrics. In Task 3, the ELFA corpus guided ChatGPT to simulate ELF conversations and engage in interactive dialogues. This aligns with Lee et al. (2025a), which developed a text-based chatbot using diverse L1 English speaker interviews. Following Chen et al. (2024), which demonstrated ChatGPT's capacity to mimic human-like pragmatic features through multiple prompting techniques, this task simulated ELF interactions

without model retraining. All prompts in Refined Model 2 were optimized before deployment.

To evaluate the GenAI-GELT module in supporting GELT, we adopted DDR, which assesses how technology tools aid pedagogy. DDR incorporates qualitative and quantitative methods, including thematic analysis, document analysis, technology reviews, and experimental analysis (Richey & Klein, 2005, 2007). It involves researchers, developers, and educators in the evaluation process. Research highlights DDR's effectiveness in analyzing ChatGPT-generated materials. For example, Choi *et al.* (2024) conducted a SWOT analysis to assess ChatGPT-generated teaching materials. Similarly, Moorhouse *et al.* (2024) highlighted the role of technological and content knowledge in evaluating ChatGPT's impact on language education, while Dai *et al.* (2024) applied technology reviews to explore GenAI's use in professional communication.

We employed a hybrid thematic analysis (Richey & Klein, 2005) combining deductive and inductive coding to comprehensively assess ChatGPT outputs. First, GE perspectives from previous research informed coding schemes, such as WE feature (Galloway & Rose, 2018; Lee *et al.*, 2025a), GE-informed assessment (Hu, 2012; Huda & Irham, 2023), and ELF strategies (Boonsuk *et al.*, 2024). Then, new themes emerged from the GenAI-generated GELT materials. To assess ChatGPT's role, we applied the affordances and constraints framework, where affordances represent ChatGPT's strengths in GELT material development, while constraints highlight its limitations (Bower & Sturman, 2015; Lee & Jeon, 2024; Lee *et al.*, 2025b). Appendix A presents a coding scheme. The first two authors independently coded and compared data, achieving Cohen's Kappa coefficient of 0.87, indicating high agreement. Discrepancies were resolved through discussions, revisiting original data as needed.

## Outcome analysis and model evaluation

Three models were used to produce GELT tasks and materials development (see GenAI-GELT, 2024, for full prompts and ChatGPT responses). Specifically, the first two models relied on ChatGPT's internal and pre-trained knowledge to create results, while the third model adopted external sources in addition to its internal knowledge to retrieve specific data to enhance model outputs.

### *Basic model: Reliance on traditional EFL approaches*

As shown in Table 3, the basic model failed to adequately represent language diversity from a GE perspective due to limitations across the four GELT dimensions. Despite being requested to incorporate elements of WE and ELF, ChatGPT's responses defaulted to NES norms, limiting its ability to represent linguistic diversity and pragmatic strategies.

In the first task, ChatGPT's output did not align with a WE perspective. For example, Figure 2 illustrates that the essay lacked core elements of Korean English, such as English borrowings (Lee, 2020), Koreanized syntactic structures (Shim, 1999), culturally embedded pragmatic strategies (Galloway & Rose, 2015), or expressions of Korean cultural identity (Lee, 2020). This result demonstrates that despite ChatGPT's extensive

**Table 3.** Affordances and constraints of basic model from GELT views.

Task and materials	Affordances on GELT	Constraints on GELT
<ul style="list-style-type: none"> <li>■ Generating essays that reflect Korean English</li> </ul>	<ul style="list-style-type: none"> <li>■ None</li> </ul>	<ul style="list-style-type: none"> <li>■ Not representing Korean English</li> <li>■ Reflecting NES features</li> </ul>
<ul style="list-style-type: none"> <li>■ Assessing and offering feedback on essays written by Korean English learners</li> </ul>	<ul style="list-style-type: none"> <li>■ None</li> </ul>	<ul style="list-style-type: none"> <li>■ Not using a GE view on essay assessment</li> <li>■ Adopting NES norms</li> </ul>
<ul style="list-style-type: none"> <li>■ Engaging in text-based ELF communication</li> </ul>	<ul style="list-style-type: none"> <li>■ None</li> </ul>	<ul style="list-style-type: none"> <li>■ Failing to engage in ELF interaction without using ELF strategies.</li> </ul>
<ul style="list-style-type: none"> <li>■ Engaging in speech-based ELF communication</li> </ul>		<ul style="list-style-type: none"> <li>■ Failing to adopt NNES speech</li> </ul>

Note. See GenAI-GELT (2024) for full prompts and responses for three models.

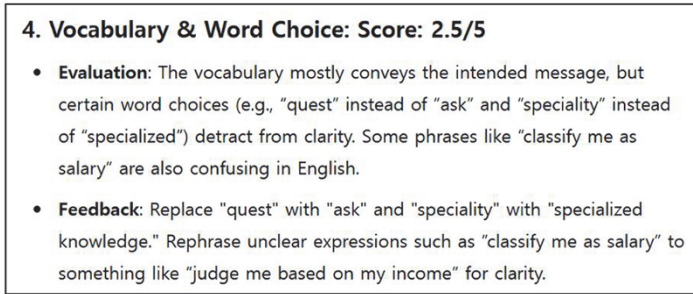
**Figure 2.** ChatGPT prompt and output for task 1 in basic model.

training on diverse language data and its sufficient knowledge of WE, ELF, and GE concepts, as evidenced in Task 2, the Basic Model was unable to produce samples reflecting English varieties. This highlights the need for refined prompting strategies, which were implemented in following models.

In Task 2, ChatGPT-produced feedback focused on standard grammar and vocabulary, overlooking linguistic diversity (WE principle), communicative effectiveness (ELF principle), and creative features of the writer's English variety. The prompt explicitly emphasized the writer's linguistic background, stating:

Evaluate an essay below written by a Korean college student and provide feedback based on World Englishes and English as a lingua franca perspectives.

This prompt aimed to acknowledge Korean English features while integrating WE and ELF perspectives, but limitations emerged. While ChatGPT's suggestions improved clarity and coherence, they followed general writing conventions rather than sociolinguistic perspectives. Consequently, the model did not apply Korean English discourse norms and stylistic features, essential for representing diverse linguistic identities within a WE-informed framework. Figure 3 illustrates how Koreanized lexical items, such as *hand phone* (instead of *mobile* or *cell phone*), were incorrectly flagged as errors



**Figure 3.** Chatgpt’s evaluation and feedback in basic model.

and replaced with standardized forms, erasing multilingual and multicultural aspects of the writer’s English. This result highlights the Basic Model’s limitation as its basic prompts fail to generate GE-informed, context-specific outputs.

ChatGPT’s facilitation of ELF interaction in Task 3 similarly fell short of representing the diversity and fluidity commonly adopted by ELF speakers (see GenAI-GELT, 2024, for full interactions between ChatGPT and a Korean English user). The prompt for this task was “Let’s do a role-play where I am a Korean college student, and you are a Chinese student studying at a British university.” Responses lacked the authenticity of ELF interaction, as it failed to incorporate Chinese English features, such as non-standard article usage, sentence-final particles, or lexical innovations. Although ChatGPT has sufficient knowledge of WE concept and extensive language training data (OpenAI, 2024), its inability to produce localized linguistic features highlights limitations in its algorithmic design, which cannot always be fully resolved through prompt-based queries. In addition, the interaction showed distraction from pragmatic strategies common to ELF communication, such as repetition for clarification, shared cultural knowledge, or code-switching to facilitate negotiation of meaning that may lead to mutual understanding among ELF speakers. These strategies were identified via the GE-informed rubrics that were used to shape the model’s outputs.

The attempt to engage in speech-based ELF communication with ChatGPT revealed significant limitations in creating realistic and diverse ELF interactions (see Task 4 in Table 3). Although the prompt explicitly requested the use of Chinese English and its diverse features, its responses did not meet these expectations: ChatGPT initially responded with “Ok. I will mimic Chinese English in our role play” but its spoken responses adhered to standard English in grammar, pronunciation, and vocabulary usage. Due to this technological limitation, the speech-based ELF communication task was excluded from refined models in the next stage.

Overall, the evaluation of ChatGPT’s performance in the basic model highlights its limitations in supporting GE principles, as its responses defaulted to standard English across four tasks. This outcome shows the limitations of single-step prompting and highlights the need for refined models capable of embracing a wide range of linguistic variations brought to English language use in global contexts.

**Table 4.** Affordances and constraints of refined model 1 for GELT.

Task and Materials	Affordance	Constraint
<ul style="list-style-type: none"> <li>■ Generating a teaching material on Korean English</li> </ul>	<ul style="list-style-type: none"> <li>■ Creating acceptable materials that reflect some features of Korean English</li> </ul>	<ul style="list-style-type: none"> <li>■ Inadequate for syntactic features of Korean English</li> </ul>
<ul style="list-style-type: none"> <li>■ Generating essays that reflect Korean English</li> </ul>		<ul style="list-style-type: none"> <li>■ Essay reflecting NES norms, not Korean English</li> </ul>
<ul style="list-style-type: none"> <li>■ Assessing and offering feedback on essays written by Korean English learners</li> </ul>	<ul style="list-style-type: none"> <li>■ None</li> </ul>	<ul style="list-style-type: none"> <li>■ Not reflecting GE-informed assessment criteria, but adopting NES norms</li> </ul>
<ul style="list-style-type: none"> <li>■ Engaging in text-based ELF communication</li> </ul>	<ul style="list-style-type: none"> <li>■ None</li> </ul>	<ul style="list-style-type: none"> <li>■ Not aligning with ELF interaction nor adopting linguistic diversity views</li> </ul>

### *Refined model 1: Lack of WE and ELF perspectives*

Although Refined Model 1 employed multi-step prompting and optimization, it still gravitated toward a singular, standardized form of English. As a result, it did not fully capture the linguistic plurality and the communicative fluidity that characterize WE and ELF perspectives. As illustrated in Table 4, the model offers minimal benefits in producing acceptable WE materials; however, it still shows limitations in fully representing linguistic diversity and fluidity.

In Task 1, ChatGPT recognized some phonological and lexical features of Korean English, including phonological, morphological features, and lexical borrowing (see GenAI-GELT, 2024). However, it displayed only superficial aspects of syntactic ones, such as topic-prominent order or selective subject omission. The essay revolved around standard English rather than incorporating specific syntactic patterns of Korean English. This created a mismatch between the identified Korean English features in the materials and their application to the essay.

For Task 2, similar to Basic Model, this refined model's evaluation and feedback on Korean English writing focused on judging errors from standard writing conventions instead of acknowledging them as variations in language use. Figure 4 shows its feedback treated local syntactic patterns, lexical choices, and repetitive structures as mistakes without considering them as reflecting the writer's Korean English background. It seems that such overreliance on NES norms in both evaluation and feedback undermined the writer's linguistic and cultural identity and voice that underpins WE.

As to Task 3, Refined Model 1 showed similar limitations as Basic Model in its capacity to produce ELF-oriented conversations. Specifically, it failed to incorporate GE-informed feedback including linguistic diversity and pragmatic strategies that are expected to emerge from text-based conversations between ChatGPT as a Chinese English speaker and a Korean speaker of English. Like the basic model, while responses in Refined Model 1 adopt simplified sentence structures, it still heavily relies on standard English grammar and vocabulary usage but lacks features of Chinese English, the result of which fell short in representing ELF interaction (see Figure 5).

Overall, Refined Model 1 made some strides beyond Basic Model but still struggled to embody WE and ELF features. Both models relied on ChatGPT's internal


**3. Feedback**

**Areas for Improvement:**

1. **Grammar:** Focus on correcting subject-verb agreement and sentence fragments (e.g., "because they think he or she wants to classify me as salary" → "because people might feel judged based on their income"). Review verb forms and pluralization rules to enhance clarity.
2. **Vocabulary:** Use precise terms to replace inaccurate phrases (e.g., "quest weight" → "ask about weight"; "Korea manner" → "Korean etiquette"). Expanding vocabulary with culturally rich expressions would also strengthen the essay.
3. **Cohesion:** Add transitions to improve flow between points (e.g., "Additionally, professional questions can be difficult during small talk because...").


Figure 4. ChatGPT’s feedback in refined model 1.

Yes I can make it. I can cook for your later. I eat Chinese mala tang many.. 'cause many mala tang restaurant near our university! you like spicy food?



Wow, Minho, you can cook? That's so cool! 😄 I will be very happy to try your ddokbokggi. Yes, I like spicy food too! Mala tang is my favorite. You know, the soup has very strong taste with Sichuan pepper and chili. It makes your mouth feel numb. 🤤 Have you tried it? I think Korean and Chinese people both love spicy food, right? Which do you think is more spicy—Korean or Chinese food? 😊

Oh...Sichuan. really? I never heard it. maybe then hm Chinese food is spicy more? haha i think.



Haha, maybe! 😄 Sichuan food is very famous in China for spicy taste. We use chili and pepper oil a lot. Some people even cry when they eat too much! 🤤

But I think Korean spicy food is more... how to say... smooth? The flavor is spicy but also sweet, like ddokbokggi. Chinese food is more... strong and heavy. You should try Sichuan hot pot one day! I can take you to a good restaurant near here. 😊 What do you think?

Figure 5. ChatGPT’s responses in a conversation with a Korean speaker in English.

knowledge, which does not support the pluralistic perspective of GE. This limitation underscores that ChatGPT’s inherent data sources and output algorithms do not sufficiently account for core GE principles, including the diversity in English usage and fluidity in ELF communication, in its responses.

*Refined model 2: Progress with remaining gaps in authenticity*

Refined Model 2, which adopted additional prompting techniques such as the RAG technique (for external knowledge sources), instructed prompting, and few-shot

**Table 5.** Affordances and constraints of refined model 2 for GELT.

Task and materials	Affordance	Constraint
■ Generating a teaching material on Korean English	■ Crafting acceptable teaching materials	■ Inadequate explanations of syntactic features
■ Generating essays that reflect Korean English		■ Essay reflecting NES norms, not Korean English
■ Assessing and offering feedback on essays written by Korean English learners	■ Acceptable evaluation and feedback from a GE view	■ Lack personalized, tailored feedback from a GE view
■ Engaging in text-based ELF communication	■ Serving as an authentic ELF communication partner	■ Not sustaining the authenticity, reverting to NES norms over time

**Figure 6.** ChatGPT-generated English essay by a Korean writer.

prompting, demonstrated both progress and persistent shortcomings. This model's responses showcased the affordances and constraints for crafting GE-informed tasks and materials (see [Table 5](#)).

Regarding the first task, the combination of multi-step prompting and RAG, which leverages external knowledge as a data source, did not significantly increase the output quality from a GE perspective. While using the Gachon Learner Corpus allowed for the identification of Korean English features, the generated 300-word essay did not satisfy the expectations for GELT materials to showcase this English variety. Like the outputs of Basic Model and Refined Model 1, it only reflected typical non-native English forms, characterized by simplified vocabulary and grammar usage (see [Figure 6](#)).

The output by Refined Model 2 for Task 2 showed notable improvements compared to Basic Model and Refined Model 1. ChatGPT provided assessments more relevant to GE principles, as its rubrics were developed using five articles on WE- and ELF-based assessment. The feedback from this model offered clear guidance for improvement from a GE perspective, such as communicative effectiveness and pragmatic strategies, while underscoring a writer's strengths in cultural relevance and creative language use. However, this model did not offer personalized feedback, as it tended to focus on standardized and abstract comments rather than a writer's individual voice, as well as unique linguistic and cultural identity. Such personalized feedback might involve

integrating dynamic prompts that account for individual learner profiles (e.g., their linguistic background), which would better reflect their unique voice and cultural identity.

For Task 3, the GenAI-GELT instruction module with the ELFA corpus initially generated improved outputs during its interaction with a user in ways to align with ELF strategies. To guide ChatGPT's responses during the ELF interaction, the model was prompted to emulate Finnish English using external data from the ELFA corpus, where Finnish speakers account for the largest proportion of participants (Mauranen *et al.*, 2010). This design choice allowed ChatGPT to reflect linguistic features commonly observed in Finnish English while interacting with a user who role-played as a speaker of Korean English. The generated interaction incorporates specific traits of Finnish English, such as hesitation markers and simplified sentence structures (see GenAI-GELT, 2024, for a conversation with ChatGPT). These elements created a seemingly authentic ELF communication style at the beginning of the conversation, reflecting the diversity and fluidity inherent in ELF interactions; however, as the dialogue progressed, the responses gradually degraded within a single chat box and shifted toward native-like English usage. Therefore, despite an initial improvement compared to earlier models, the output ultimately fell short in sustaining linguistic diversity and authentic representation of ELF communication throughout the interaction. This observation suggests that ChatGPT's internal algorithm may favor standardization over time, leading to a gradual drift toward dominant linguistic forms. The result highlights the inherent challenge of maintaining diverse linguistic representations without continuous prompt-based interventions.

In short, Refined Model 2 shows noticeable improvements over previous models, particularly in Task 2 and Task 3, due to the use of multi-step prompting and external knowledge sources. Regarding its potential to support GE principles, Task 2 benefited from a more structured evaluation in accordance with GE principles, while Task 3 outputs initially demonstrated elements of authentic ELF communication. However, its limitations were also identified, such as standardized feedback in Task 2 and a gradual shift toward NES norms in Task 3. These results highlight the progress achieved by Refined Model 2 but also underscore the need for further refinement to fully support GE-focused tasks.

## Discussion and reflections

This article explored the potential of ChatGPT, an LLM-based GenAI tool, in supporting ELT from a GELT perspective, specifically focusing on its ability to design GE-informed ELT tasks and materials. In response to the two bias sources, including data representation and statistical probabilities inherent in algorithmic design, this study aimed to improve the base model by employing prompt engineering techniques and ELFA corpus data as an external knowledge source. The analysis shows that the first refined model displayed minimal improvements, indicating its limited effectiveness in generating GE-informed outputs. In contrast, Refined Model 2 demonstrated significant enhancements by producing acceptable outputs that better represent linguistic diversity and support ELF interactions. These improvements were achieved by addressing biases, despite certain limitations in data actualization and output authenticity.

**Table 6.** Three models' affordances and constraints on GELT.

Model	Affordances	Constraints
Basic Model	<ul style="list-style-type: none"> <li>■ None</li> </ul>	<ul style="list-style-type: none"> <li>■ Sticking to NES norms in all tasks</li> </ul>
Refined Model 1	<ul style="list-style-type: none"> <li>■ Creating acceptable WE-informed materials</li> </ul>	<ul style="list-style-type: none"> <li>■ Sticking to NES norms in all the other tasks</li> </ul>
Refined Model 2	<ul style="list-style-type: none"> <li>■ Creating acceptable WE-informed materials</li> <li>■ Acceptable evaluation and feedback from a GE view</li> <li>■ Serving as an authentic ELF communication partner</li> </ul>	<ul style="list-style-type: none"> <li>■ Reflecting standard English in a generated essay</li> <li>■ Lack personalized, tailored feedback</li> <li>■ Not sustaining the authenticity, reverting to NES norms over time</li> </ul>

Table 6 provides a summary of the three models' affordances and constraints in supporting GE-informed ELT tasks and materials.

Refined Model 1's minimal improvements reveal the fundamental limitations inherent in ChatGPT's internal knowledge despite utilizing prompt engineering techniques. For example, this model's tendency to default to NES norms, rather than representing the linguistic diversity common to GE usage, highlights the critical influence of training data biases. This indicates that the overrepresentation of dominant varieties of English in the original training datasets default can marginalize underrepresented varieties, thereby resulting in outputs that lack authenticity in ways to capture the pluralistic realities of English used in intercultural communication (Dai et al., 2024; Hovy & Prabhumoye, 2021; Jeon et al., 2025). From an ELF perspective, this limitation is particularly evident in the absence of lingua-culturally diverse repertoire within the basic model's datasets (Dai et al., 2024; Jeon et al., 2025; Payne et al., 2024), which restricts its ability to generate the context-sensitive forms necessary for effective meaning negotiation in ELF interactions (Boonsuk et al., 2024; Lee et al., 2025a). Without such data resources as linguistic repertoires, LLMs struggle to articulate linguistic forms aligned with the dynamic, adaptive nature of ELF communication, where the local construction of linguistic norms among interlocutors depends on contextual flexibility (Mu et al., 2023; Seidlhofer, 2011). Therefore, the limitations identified in Refined Model 1 showcase that the critical role of internal data sources entails the challenge of how GenAI agents interact with ELF users, as the inherent biases in GenAI datasets may inadvertently perpetuate monolingual ideologies. This seems to occur by statistically favoring dominant linguistic forms and marginalizing diverse language practices, which can be interpreted as deficient due to the model's reliance on probabilistic output generation (Payne et al., 2024). Put simply, if teachers use GenAI to produce "authentic" materials via simple prompting of the internal LLM, there is a danger that it will produce highly sanitized materials that are not representative of how English is used globally. While this is most directly applicable to GELT-oriented teachers, similar challenges could arise in other ELT contexts where diverse linguistic inputs and non-standard varieties of English play a critical role in learning outcomes. For this, RAG datasets can enhance ChatGPT's inclusivity (Sahoo et al., 2024).

Refined Model 2 that adopted external datasets and advanced prompts showed a significant improvement in the outputs' authenticity. In Task 3, this model specifically demonstrated flexible ELF strategies using the ELFA corpus, including hesitation markers and simplified structures that are reflective of Finnish English. These outputs marked an adequate shift from NES norms observed in earlier models. This progress highlights the importance of utilizing high-quality external data sources in refining GenAI outputs, as the diversity and reliability of such datasets may directly influence GenAI's capacity to demonstrate the pluralistic and fluid nature of GE (Dai et al., 2024; Navigli et al., 2023). The significance of using quality data sources extends beyond simply diversifying linguistic representation; it can also enhance the contextual relevance of GenAI outputs by mitigating biases. Payne et al. (2024) and Jenks (2025) emphasize the role of lingua-culturally rich data in addressing the inherent limitations of standardized training datasets, which often reinforce monolingual ideologies. For this purpose, Zhu et al. (2025) emphasizes the necessity of collaboration among technologists, linguists, and educators to curate and integrate datasets that reflect real-world linguistic diversity. This approach can ensure that GenAI not only generates authentic outputs but also aligns with GELT principles by respecting cultural identities and fostering equitable communication practices (Lee et al., 2025a; Rose & Galloway, 2019).

Refined Model 2's improvement was not exclusively due to adopting authentic external datasets but also resulted from using multiple prompting techniques. The comparison among the three models shows that this way of prompt usage played a critical role in guiding ChatGPT's responses toward GELT-specific outputs. Research shows that combining multiple prompting techniques may enhance the contextual relevance and adaptability of LLM outputs while mitigating biases toward NES norms (Sahoo et al., 2024). In this study, RAG enabled ChatGPT to retrieve task-specific data from external sources, while instructed and few-shot prompting helped refine its responses to better reflect diverse linguistic features. The combined use of these techniques is particularly significant for the development of GELT-oriented ELT materials, as it offers a practical method for integrating linguistic diversity into classroom instruction without requiring extensive technical retraining of the model (Lee et al., 2024b; Rose et al., 2021). This further indicates that English teachers, utilizing multiple prompting techniques, can create contextually valid instructional materials that expose learners to diverse English varieties and authentic language use in ELF communication contexts (Lo, 2025).

Despite using external corpus data that includes diverse English varieties and ELF episodes, Refined Model 2 still displayed constraints in contextualizing and personalizing its outputs. As observed in Task 1, the model struggled to generate an essay that reflects a target variety and finally defaulted to NES forms. In Tasks 2 and 3, while the model demonstrated some adaptability, its responses lacked the specificity that represents pragmatic strategies common to ELF communication. Particularly, Refined Model 2's responses in Task 3, which initially aligned with ELF interactions in the corpus but eventually reverted to native norms, reflects its tendency toward standardization and dominant forms despite providing ELF corpus data. This issue underscores the challenges inherent in the current ChatGPT model's algorithm design and statistical probabilities, which favor predictable outputs over dynamic interaction in ELF (Bender & Koller, 2020; Jenkins, 2015). The fluidity and hybridity of ELF

make it inherently challenging to develop a stable, one-size-fits-all model that fully attests to its dynamic nature. However, gradual improvements may be possible through multi-step prompt strategies and external data integration, allowing for more context-specific adaptability in local ELF contexts (Sahoo et al., 2024). To address the inherent challenges, in other words, further customization is required, such as incorporating additional ELF-based assessment criteria into the prompts. Strategies may be needed such as refreshing the model's prompts mid-conversation or using multi-step prompts to maintain linguistic diversity throughout the interaction.

These findings raise critical questions about whether GenAI outputs can authentically represent the identities of specific ELF users or engage in identity-building dialogues (Galloway & Rose, 2015; Seidlhofer, 2011). The success of ELF communication may depend on the mutual construction of local norms within a community of practice (CoP), where interlocutors actively invest their linguistic and cultural resources in negotiating meaning and constructing identity (Jenkins, 2015; Mu et al., 2023). However, it does not seem that ChatGPT's outputs in its conversation with a human speaker reflect such dynamic co-construction of a CoP; but instead, it may generate static linguistic patterns in its responses. Thus, the authenticity of AI-generated ELF interactions remains debatable, as Lee et al. (2025a) classify such verbal exchanges as "ELF-like interactions" and recommend their usage for raising GE awareness.

## Implications

The analyses suggest some pedagogical implications for the effective use of GenAI for GELT, as well as ELT more broadly. We have grouped these recommendations under three themes: (1) GenAI-teacher collaboration in iterative design processes, (2) the significance of teachers' professional knowledge in developing GELT tasks with GenAI tools, and (3) teachers' agentive role in orchestrating complementary resources to address biases and enhance inclusivity. These implications highlight that GenAI's ELT success depends on teachers.

First, the findings underscore the importance of teachers' collaboration with GenAI, particularly in utilizing prompt engineering and selecting external sources to maximize GenAI's performance (Jeon & Lee, 2023). This collaboration involves more than solely obtaining specific outputs; rather, it represents an iterative process of co-creation where teachers and GenAI mutually refine and enhance the quality of generated content (Choi et al., 2024). Teachers' roles go beyond simply adapting to AI outputs – they actively shape and direct the process by making pedagogically sound decisions, interpreting AI-generated suggestions, and integrating them into lesson plans (Jeon & Lee, 2023). Specifically, this process of AI-teacher collaboration requires teachers to design prompts, analyze the responses, and modify inputs in an iterative manner for specific pedagogical goals (Lee et al., 2024a). Jeon and Lee (2023) particularly emphasize language teachers' ability to interpret and adjust ChatGPT outputs in ways to satisfy pedagogical needs and enhance student engagement. The effective use of GenAI for pedagogical purposes is closely tied to teachers' technological knowledge to customize GenAI tools in specific educational contexts in which teachers need training in prompting (Moorhouse, 2024).

Second, teachers' professional knowledge is crucial for designing appropriate tasks and materials with GenAI tools, as well as to evaluate the appropriateness of outputs for classroom use (Lee *et al.*, 2025b). Promoting GELT-specific content knowledge among teachers could be facilitated through teacher education programs, professional development workshops, and continuous training sessions designed to address the integration of GenAI tools into ELT curricula (Lee & Jeon, 2024; Lo, 2025). The findings in the present study further indicate that teachers' GELT-specific content knowledge is essential for integrating external datasets, such as ELF and WE corpora, to address linguistic biases inherent in GenAI's internal database and to produce GE-informed authentic outputs. In this study, the inclusion of ELF corpora significantly improved the quality of ChatGPT outputs, thereby enabling ChatGPT to better support GELT. This highlights the importance of teachers' active use of their expertise in selecting and integrating external data sources to maximize the pedagogical potential of GenAI (Jeon & Lee, 2023). Beyond the scope of GELT, this professional knowledge would also be essential in using GenAI for other areas of ELT, such as business writing, EAP, and ESP, where teachers' professional knowledge is needed to ensure GenAI outputs are suited for the curriculum and not accepted at face-value.

Finally, the persistent biases in LLM-based GenAI outputs, despite the use of advanced prompt engineering techniques and ELFA corpus data in our study, underscore the need for teachers to orchestrate complementary resources to maximize pedagogical effectiveness (Jeon & Lee, 2023; Moorhouse, 2024). Specifically, earlier GELT methods must be used in tandem with GenAI tools to address the limitations from two primary GenAI biases (Lee *et al.*, 2025a; Rose *et al.*, 2021). For example, teachers can incorporate more role-play diversity: by expanding role-play scenarios to include more diverse ELF contexts (e.g., interactions between speakers with other localized English varieties), they can better demonstrate the fluid and adaptive nature of ELF. For more diverse curriculum development, teachers can incorporate author profiles into written ELF corpora to account for ELF users' linguistic identity and cultural backgrounds in GE-informed written assessments (Rose & Galloway, 2019). Also, integrating multimedia resources with GenAI usage can create a holistic learning environment that supports ELF communication (Lee *et al.*, 2025a).

### Concluding remarks and future directions

This study shows ChatGPT's potential for GELT by mitigating biases and enhancing its performance. Through iterative refinement processes involving prompt engineering and the integration of external datasets, such as ELF corpora, refined models demonstrated significant improvements in producing linguistically diverse and culturally authentic outputs. Standard language biases persist, highlighting the teacher's role. These findings highlight teacher agency in GenAI use. For this purpose, teacher professional development programs should focus on equipping teachers with GELT-specific knowledge, including the ability to select and integrate diverse language datasets and design GE-informed tasks. Training helps teachers integrate GenAI into ELT (Jeon & Lee, 2023).

This study can extend to future research topics at the intersection between GenAI and GELT. First, future research should explore the multifaceted dimensions of teacher

agency in utilizing GenAI tools for GELT, including how teachers develop and apply prompt engineering techniques, curate external datasets, and integrate GenAI outputs into diverse educational contexts (Lee & Jeon, 2024; Moorhouse et al., 2024). Second, the use of GenAI for GE-informed assessment requires further investigation, particularly in developing corpora that include actual evaluation and feedback to enhance the authenticity of GenAI outputs for GELT (Rose & Galloway, 2019; Tsagari et al., 2023). To achieve this, GELT-informed assessment practices may prioritize evaluating communicative effectiveness and intercultural adaptability. As Nakamura (2020) and Tsagari et al. (2023) note, incorporating non-native speaker-driven corpora and dynamic rubrics can help address the inherent biases of the current assessment practices in ELT. This can be achieved through flexible assessments that emphasize mutual intelligibility, strategic competence, and context-specific language use (Rose & Galloway, 2019). Future research can explore the role of interactive GenAI-driven assessments, enabling learners to engage dynamically with diverse linguistic data, thereby reflecting authentic English use across contexts (Lee & Jeon, 2023). Third, future studies may explore curriculum innovation by adopting GenAI to design GE-informed materials and activities. This can focus on designing dynamic, learner-centered curricula that promote intercultural communication skills, awareness of linguistic diversity, and critical engagement in GE contexts (Rose et al., 2021). Fourth, research should examine the role of GenAI in ELF interaction training to explore how this technology can provide authentic conversational practices, simulate diverse linguistic and cultural contexts, and reduce communication anxiety (Lee et al., 2025b). Fifth, the findings in this study imply that there is a critical need to address identity construction and the use of lingua-cultural resources in GenAI-mediated ELF communication. Thus, future research can investigate how human ELF users engage in communication with GenAI chatbots, particularly focusing on CoP to provide insights into the dynamic interplay between ELF users' linguistic identities and GenAI's responses, as well as the authenticity of GenAI-mediated ELF interactions. Further research to inform the appropriate use of GenAI in ELT will help to harness its full power to develop materials, give feedback, and provide interactive experiences, while avoiding the pitfalls of producing sanitized and standard language forms that are not globally representative of the language in use in a diverse range of authentic settings. Finally, while this study focuses on making GenAI tools usable for teachers in classroom-based ELT contexts, we recognize the potential of collaboration between ELT professionals and AI developers in addressing some of the identified limitations in this study (Chang et al., 2022). Future research could develop task-specific datasets and AI agents to improve linguistic diversity in AI outputs. Although this study does not explore algorithmic redesign, such collaborative efforts could complement teacher-driven prompt engineering and foster more inclusive GE-oriented instructional materials. Future research could further investigate the potential of retraining GenAI models via Application Programming Interfaces (APIs) to enhance the incorporation of external, corpus-based knowledge and address limitations inherent in prompt-based approaches (Kim & Lu, 2024; Lee et al., 2024b). APIs can facilitate the direct transfer of external data and instructions to the model, enabling fine-tuning of its internal parameters for improved task-specific accuracy (OpenAI, 2024). In this way, API-driven retraining offers a relative advantage over prompt engineering, which

externally guides model outputs only. This internal adjustment has the potential to create more contextually relevant GELT materials while maintaining linguistic diversity across a broader range of tasks (Lee *et al.*, 2024b).

This study has several limitations. First, as the study focused on the Korean variety of English, the findings may not be directly generalizable to other localized English varieties without further contextual adaptation. Future research could explore how the methodology and findings in this study can extend to diverse linguistic contexts by incorporating additional local English varieties and analyzing their pedagogical implications. Second, while Refined Model 2 showed improvements, it still struggled to fully sustain GELT-aligned materials and tasks, occasionally reverting to NES norms. This highlights the limitations of prompt engineering alone. Future research could develop an enhanced GenAI-GELT Instructional Module using API-based fine-tuning to better support linguistic diversity and GELT principles.

## References

- Abdelhalim, S. M. (2024). Using ChatGPT to promote research competency: English as a Foreign Language undergraduates' perceptions and practices across varied metacognitive awareness levels. *Journal of Computer Assisted Learning*, 40(3), 1261–1275. <https://doi.org/10.1111/jcal.12948>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–5198. <https://doi.org/10.18653/v1/2020.acl-main.463>
- Boonsuk, Y., Wasoh, F., & Ambele, E. A. (2024). Global Englishes pedagogical activities for English-as-a-foreign language settings and beyond: Understanding Thai teachers' practices. *RELC Journal*, 55(2), 379–393. <https://doi.org/10.1177/00336882221112193>
- Bower, M., & Sturman, D. (2015). What are the educational affordances of wearable technologies? *Computers and Education*, 88, 343–353. <https://doi.org/10.1016/j.compedu.2015.07.013>
- Brandt, A., & Hazel, S. (2025). Towards interculturally adaptive conversational AI. *Applied Linguistics Review*, 16(2), 775–786. <https://doi.org/10.1515/applirev-2024-0187>
- Carlstrom, B., & Price, N. (2014). *The Gachon learner corpus* [Corpus data]. <http://koreanlearnercorpusblog.blogspot.kr/p/corpus.html>
- Chang, Y., Lee, S., Wong, S. F., & Jeong, S. (2022). AI-powered learning application use and gratification: An integrative model. *Information Technology and People*, 35(7), 2115–2139. <https://doi.org/10.1108/ITP-09-2020-0632>
- Chen, X., Li, J., & Ye, Y. (2024). A feasibility study for the application of AI-generated conversations in pragmatic analysis. *Journal of Pragmatics*, 223, 14–30. <https://doi.org/10.1016/j.pragma.2024.01.003>
- Choi, G. W., Kim, S. H., Lee, D., & Moon, J. (2024). Utilizing generative AI for instructional design: Exploring strengths, weaknesses, opportunities, and threats. *TechTrends*, 68(4), 832–844. <https://doi.org/10.1007/s11528-024-00967-w>
- Cooke, S. (2020). Assessing real-world use of English as a lingua franca (ELF): A validity argument. *VNU Journal of Foreign Studies*, 36(4), 47–62. <https://doi.org/10.25073/2525-2445/vnufs.4574>
- Dai, D. W., Suzuki, S., & Chen, G. (2024). Generative AI for professional communication training in intercultural contexts: Where are we now and where are we heading? *Applied Linguistics Review*, 16(2), 763–774. <https://doi.org/10.1515/applirev-2024-0184>
- ELFA. (2008). *The Corpus of English as a lingua franca in academic settings* (Director: Anna Mauranen) [Corpus data]. University of Helsinki. <https://www.helsinki.fi/en/researchgroups/english-as-a-lingua-franca-in-academic-settings/research/elfa-corpus>
- Galloway, N., & Rose, H. (2015). *Introducing Global Englishes*. Routledge.

- Galloway, N., & Rose, H. (2018). Incorporating Global Englishes into the ELT classroom. *ELT Journal*, 72(1), 3–14. <https://doi.org/10.1093/elt/ccx010>
- GenAI-GELT. (2024). *GenAI-GELT instructional module* (1–20) [GenAI Output]. <https://docs.google.com/document/d/1xItC7f02UN3nJXBfj8JV9neuFSNforzAyyLjZO6NjA/edit?usp=sharing>
- 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>
- Hovy, D., & Prabhumoye, S. (2021). Five sources of bias in natural language processing. *Language and Linguistics Compass*, 15(8), e12432. Article. <https://doi.org/10.1111/lnc3.12432>
- Hu, G. (2012). Assessing English as an International language. In L. Alsagoff, S. Lee McKay, G. Hu & W. A. Renandya (Eds.), *Principles and practices for teaching English as an International Language*. Routledge, pp. 123–143. <https://doi.org/10.4324/9780203819159>
- Huda, M., & Irham, I. (2023). Contesting the nativelikeness norms of productive skill assessment in the peripheral ELT practice: ELF and world Englishes perspectives. *MEXTESOL Journal*, 47(2), 109.
- Jenkins, J. (2015). Repositioning English and multilingualism in English as a Lingua Franca. *Englishes in Practice*, 2(3), 49–85. <https://doi.org/10.1515/eip-2015-0003>
- 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>
- Jeon, J., Li, W., Tai, K. W., & Lee, S. (2025). Generative AI and its dilemmas: exploring AI from a translanguaging perspective. *Applied Linguistics*, 46(4), 709–717. <https://doi.org/10.1093/applin/amaf049>
- Karataş, F., Abedi, F. Y., Ozek Gunyel, F., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education and Information Technologies*, 29(15), 19343–19366. <https://doi.org/10.1007/s10639-024-12574-6>
- Kim, M., & Lu, X. (2024). Exploring the potential of using ChatGPT for rhetorical move-step analysis: The impact of prompt refinement, few-shot learning, and fine-tuning. *Journal of English for Academic Purposes*, 71, 101422. <https://doi.org/10.1016/j.jeap.2024.101422>
- Kumaravadivelu, B. (2012). Individual identity, cultural globalization, and teaching English as an international language. In L. Alsagoff, S. L. McKay, G. Hu, & W. A. Renandya (Eds.), *Principles and practices for teaching English as an international language* (pp. 1–19). Routledge.
- Lee, J. H., Shin, D., & Noh, W. (2023). Artificial intelligence-based content generator technology for young English-as-a-foreign-language learners' reading enjoyment. *RELC Journal*, 54(2), 508–516. <https://doi.org/10.1177/00336882231165060>
- Lee, S. (2020). Attitudes toward English borrowings in South Korea: A comparative study of university professors and primary/secondary teachers of English. *Asian Englishes*, 22(3), 238–256. <https://doi.org/10.1080/13488678.2019.1684622>
- Lee, S., & Jeon, J. (2023). Addressing automatic speech recognition for ELT from the global Englishes perspective. *ELT Journal*, 77(4), 435–444. <https://doi.org/10.1093/elt/ccad038>
- Lee, S., & Jeon, J. (2024). Teacher agency and ICT affordances in classroom-based language assessment: The return to face-to-face classes after online teaching. *System*, 121, 103218. <https://doi.org/10.1016/j.system.2023.103218>
- Lee, S., Jeon, J., & Choe, H. (2025a). Enhancing pre-service teachers' Global Englishes awareness with technology: A focus on AI chatbots in 3D metaverse environments. *TESOL Quarterly*, 59(1), 49–74. <https://doi.org/10.1002/tesq.3300>
- Lee, S., Jeon, J., & Choe, H. (2025b). Generative AI (GenAI) and pre-service teacher agency in ELT. *ELT Journal*, 79(2), 287–296. <https://doi.org/10.1093/elt/ccaf005>
- Lee, U., Jeon, M., Lee, Y., Byun, G., Son, Y., Shin, J., Ko, H., & Kim, H. (2024b). LLaVA-docent: Instruction tuning with multimodal large language model to support art appreciation education. *Computers and Education: Artificial Intelligence*, 7, 100297. <https://doi.org/10.1016/j.caeai.2024.100297>
- Lee, U., Jung, H., Jeon, Y., Sohn, Y., Hwang, W., Moon, J., & Kim, H. (2024a). Few-shot is enough: Exploring ChatGPT prompt engineering method for automatic question generation in English education. *Education and Information Technologies*, 29(9), 11483–11515. <https://doi.org/10.1007/s10639-023-12249-8>

- Lo, A. W. T. (2025). The educational affordances and challenges of generative AI in Global Englishes-oriented materials development and implementation: A critical ecological perspective. *System*, 130, 103610. <https://doi.org/10.1016/j.system.2025.103610>
- Mauranen, A., Hynninen, N., & Ranta, E. (2010). English as an academic lingua franca: The ELFA project. *English for Specific Purposes*, 29(3), 183–190. <https://doi.org/10.1016/j.esp.2009.10.001>
- 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., Wan, Y., Wu, C., Kohnke, L., Ho, T. Y., & Kwong, T. (2024). Developing language teachers' professional generative AI competence: An intervention study in an initial language teacher education course. *System*, 125, 103399. <https://doi.org/10.1016/j.system.2024.103399>
- Mu, Y., Lee, S., & Choe, H. (2023). Factors influencing English as a lingua franca communication: A case of an international university in China. *System*, 116, 103075. <https://doi.org/10.1016/j.system.2023.103075>
- Nakamura, Y. (2020). Assessing English as a lingua franca in an academic context: An ELF-aware approach. *Keio University Hiyoshi Review: Language, Culture and Communication*, 52, 145–155.
- Navigli, R., Conia, S., & Ross, B. (2023). Biases in large language models: Origins, inventory, and discussion. *Journal of Data and Information Quality*, 15(2), 1–21. <https://doi.org/10.1145/3597307>
- OpenAI. (2024). Hello GPT-4o. *OpenAI*. <https://openai.com/index/hello-gpt-4o/>
- Payne, A. L., Austin, T., & Clemons, A. M. (2024). Beyond the front yard: The dehumanizing message of accent-filtering technology. *Applied Linguistics*, 45(3), 553–560. <https://doi.org/10.1093/applin/amae002>
- Richey, R. C., & Klein, J. D. (2005). Developmental research methods: Creating knowledge from instructional design and development practice. *Journal of Computing in Higher Education*, 16(2), 23–38. <https://doi.org/10.1007/BF02961473>
- Richey, R. C., & Klein, J. D. (2007). *Design and development research: Methods, strategies, and issues*. Routledge. <https://doi.org/10.4324/9780203826034>
- Rose, H., & Galloway, N. (2019). *Global Englishes for language teaching*. Cambridge University Press.
- Rose, H., & McKinley, J. (2025). Global Englishes and TESOL: An editorial introduction to innovating research and practice. *TESOL Quarterly*, 59(1), 5–23. <https://doi.org/10.1002/tesq.3373>
- Rose, H., McKinley, J., & Galloway, N. (2021). Global Englishes and language teaching: A review of pedagogical research. *Language Teaching*, 54(2), 157–189. <https://doi.org/10.1017/S0261444820000518>
- Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2024). *A systematic survey of prompt engineering in large language models: Techniques and applications*. arXiv. <http://arxiv.org/abs/2402.07927>
- Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford University Press.
- Shim, R. J. (1999). Codified Korean English: Process, characteristics and consequence. *World Englishes*, 18(2), 247–258. <https://doi.org/10.1111/1467-971X.00137>
- Tsagari, D., Reed, K., & Lopriore, L. (2023). Teacher beliefs and practices of language assessment in the context of English as a lingua franca (ELF): Insights from a CPD course. *Languages*, 8(1), 58. <https://doi.org/10.3390/languages8010058>
- Van Horn, K. (2024). ChatGPT in English language learning: Exploring perceptions and promoting autonomy in a university EFL context. *Tesl-ej*, 28(1), 1. <https://doi.org/10.55593/ej.28109a8>
- Vatsal, S., & Dubey, H. (2024). *A survey of prompt engineering methods in large language models for different NLP tasks*. arXiv. <http://arxiv.org/abs/2407.12994>
- 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>
- Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages*, 8(3), 212. <https://doi.org/10.3390/languages8030212>
- Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90–112. <https://doi.org/10.1111/bjet.13370>
- Zhu, H., Dai, D. W., Brandt, A., Chen, G., Ferri, G., Hazel, S., Jenks, C., Jones, R., O'Regan, J., & Suzuki, S. (2025). Exploring AI for intercultural communication: Open conversation. *Applied Linguistics Review*, 16(2), 809–824. <https://doi.org/10.1515/applirev-2024-0186>

## Appendix A. Coding scheme for thematic analysis

Task topic	Theme	Code
Task 1. WE feature	1. Diverse English varieties	1.1. Integrating inner circle varieties 1.2. Integrating outer circle varieties 1.3. Integrating expanding circle varieties
	2. Valuing non-native norms	2.1. Non-native speaker communication strategies 2.2. Non-native interactional norms 2.3. Successful communication of non-native speakers
	3. Linguistic diversity in ELT	3.1. L1 usage in English instruction 3.2. Recognizing multilingual competence as an asset 3.3. Incorporating instances of linguistic hybridity
	4. Challenging standard English ideology	4.1. Exposure to diverse English varieties 4.2. Critiquing native-speakerism 4.3. Authentic English use in multilingual contexts
Task 2. GE-informed assessment	1. Intelligibility & comprehensibility	1.1. Intelligibility with non-native speakers 1.2. Tolerance for linguistic creativity and variation 1.3. Focus on meaning over correctness
	2. Language use (structure)	2.1. Flexibility on non-standard forms 2.2. Allowing errors without impeding communication 2.3. Allowing multilingual strategies
	3. Language use (vocabulary)	3.1. Contextually appropriate lexical diversity 3.2. Vocabulary creativity, including hybrid forms 3.3. Accepting local word choices
	4. Cohesion and organization	4.1. Multiple (e.g., non-linear) coherence forms 4.2. Flexible on diverse transition & discourse markers 4.3. Stressing flow over conventional organization
	5. Cultural relevance	5.1. Culturally specific references 5.2. Valuing non-native perspectives and local contexts 5.3. Accepting culturally diverse rhetorical styles

(Continued)

(Continued.)

Task topic	Theme	Code
	6. Communication strategies	6.1. Effective use of accommodation strategies 6.2. Simplification & strategic code-switching 6.3. Boosting understanding through paraphrasing
Task 3. ELF strategy	1. Negotiation and accommodation	1.1. Clarification requests 1.2. Paraphrasing and reformulations 1.3. Aligning speech patterns to interlocutor needs
	2. Use of multilingual resources	2.1. Code-switching between languages 2.2. Translanguaging (fluid switching/blending) 2.3. Borrowing and hybridization of linguistic forms
	3. Expression of identity and solidarity	3.1. Projecting personal and cultural identities 3.2. Using linguistic forms to signal group membership 3.3. Emphasizing shared experiences or references
	4. Managing conversations	4.1. Effective turn-taking and sequencing 4.2. Managing pauses 4.3. Using culturally relevant discourse markers
	5. Managing repair and error	5.1. Clarification requests to resolve ambiguity 5.2. Repetition/reformulation of misunderstood phrases 5.3. Strategic rephrasing to handle misunderstandings