

AI Tools and EFL Identity: A Critical Discourse Analysis of Gender and Cultural Representation in AI-Generated Reading Materials for Gulf University Students

Adnan Mohammed Bataineh¹, Kashif Ali Sabiri¹, Ebrahim Douzandeh¹

¹ University of Buraimi, Oman

Correspondence: Ebrahim Douzandeh, University of Buraimi, Oman. P. O. Box: 890.

Received: May 12, 2026

Accepted: September 2, 2026

Online Published: October 16, 2026

doi:10.5430/wjel.v17n1p160

URL: <https://doi.org/10.5430/wjel.v17n1p160>

Abstract

The incorporation of artificial intelligence (AI) tools into English as a Foreign Language (EFL) classroom has revolutionized the production of teaching content. However, the ideological implications of AI-generated content have been markedly under-researched in non-Western contexts. This study employs Fairclough's three-dimensional framework of Critical Discourse Analysis (CDA) on a corpus of 45 AI-generated EFL reading texts produced by the ChatGPT-4o, Google Gemini 1.5 Pro, and Microsoft Copilot in response to five standardized instructional prompts. It aims to investigate how gender and Arab cultural identity is discursively constructed in materials targeting Gulf university EFL learners. Analysis of the findings show that the female professional is consistently represented through discourse of relational and communal achievement while the single male professional is coded through individual enterprise and resilience. Leadership discourse universalizes the individual while entirely erasing gender and cultural identity across the nine texts. Furthermore, a Western nuclear family development narrative renders the Arab family invisible. Finally, Microsoft Copilot shows the lowest cultural specificity and the highest degree of textual cloning across platforms. The findings further affirm that L2 writing created by these AI models replicates Western liberal ideological formations, thus marginalizing the cultural and gender identities of the Arab learners. The research makes pedagogical, technological, and policy recommendations for the culturally responsive use of AI tools in higher education in the Gulf.

Keywords: critical discourse analysis; AI-generated EFL materials; gender representation; Arab cultural identity; Gulf higher education

1. Introduction

1.1 AI-Generated Content in EFL Education

Whose gender identities and whose cultural identities are constructed for Arab EFL learners when the reading materials they study are authored not by teachers but by AI systems? This is the central identity question this study addresses. The emergence of AI tools as content-generation platforms in English as a Foreign Language (EFL) higher education raises pressing questions about whose knowledge, identities, and cultural values are built into the materials they produce. Curriculum designers and teachers can now create reading passages, discussion prompts, grammar tasks, and assessment tasks in seconds using platforms like ChatGPT, Google Gemini, and Microsoft Copilot. Within the context of higher education, in particular, AI-generated materials are getting increasingly supplementary, and in certain cases, even replacing traditionally authored resources, with little to no critical analysis of the gendered and culturally located identities and assumptions inherent to them (Bender et al., 2021; Ge et al., 2025; Huang et al., 2025).

The use of generative AI in language teaching has been recognized by recent studies to have a deep-rooted tension at its core. Large language models (LLMs) are trained in excess on English language, Western-centric data that is selected and processed by overwhelmingly homogeneous workforces (Dodge et al. 2021; Mushtaq et al., 2026; Fenech-Borg et al., 2025; Daryani et al., 2026). Following this, AI-generated educational texts, however, are not the result of neutral, objective processes; they hold hidden assumptions about whose knowledge is valued, whose social experiences are normal, and whose cultural and gender identities are centered or peripheral in the experience of the learner of the content.

In this study, EFL identity is treated as the overarching construct under investigation: the discursive positioning of Arab EFL learners as gendered and culturally located subjects within the reading materials through which they learn English. Gender representation and cultural representation, the two empirical foci named in the title, are the two principal dimensions through which this identity construction is examined, while ideology and bias are treated as the underlying mechanisms that produce it, rather than as separate objects of study in their own right. Framing the inquiry in terms of identity therefore situates the analysis of AI, ideology, and representation within a single coherent aim: to determine what version of the Arab EFL learner, as a gendered and cultural subject, is implicitly authorized or excluded by AI-generated reading materials.

1.2 Discourse, Power, and AI-Generated Texts

Critical discourse analysis (CDA) is an investigation into the relationship between language, power, and ideology. It is a theoretical, analytical, and methodological framework which sees language as a social practice whose properties both shape and are shaped by power relations (Fairclough, 2023). According to Fairclough's three-dimensional model, any example of text is to be analyzed simultaneously at the level of text (lexis, grammar, modality), at the level of discursive practice (the way the text is produced, circulated, consumed), and at the level of sociocultural practice (the ideological and institutional conditions of discourse). According to CDA, texts created by machines and humans are not ideologically neutral; they reflect the life conditions of their production while reinforcing those conditions (Fairclough, 2023). Blodgett et al. (2020) show that "neutral" language technologies inherently reflect social inequalities, and that they do so in a systematic way, typically with measurable impacts on groups that do not share the dominant social power. Applying CDA to the AI-generated educational content is not only a methodological requirement but also a critical necessity to ensure the AI-mediated pedagogy does not work as a tool of cultural and epistemic exclusion.

While these studies provide important insights, each has a scope that leaves the present research question unaddressed. Ahmed and Mahmood (2024) examined ChatGPT's argumentative responses to contested political knowledge questions, not the instructional reading materials that EFL learners are assigned to study, and their analysis targeted ideological bias in propositional content rather than gender or cultural representation. Ting et al. (2025) applied Fairclough's model to AI-generated feedback on student writing, a genre of AI output that is evaluative and addressed to an individual learner's text, rather than to the expository reading passages, discussion prompts, and comprehension materials that constitute the primary content Gulf EFL curricula deliver to learners at scale. Neither study conducted a transitivity-level or lexical-collocation-level analysis of how gendered or culturally marked protagonists are constructed, and neither examined a corpus specifically designed to compare outputs across multiple LLM platforms under controlled, repeated prompting. Consequently, a clear methodological gap remains to date, the AI-generated EFL reading materials through which learners encounter and internalize the AI's hidden assumptions have not been studied using a systematic, corpus-based CDA approach for their gendered and cultural representations, particularly in the Global South and the Arab world. To date, the AI-generated EFL reading materials through which learners encounter and internalize the AI's hidden assumptions have not been studied in a systematic way for their gendered and cultural representations. This is particularly the case in the Global South and the Arab world.

1.3 Gulf University Students as an Underrepresented Population

The largest but least researched group in the field of applied linguistics is the Arab EFL learners in the Gulf Cooperation Council (GCC) countries, including Oman. These learners are positioned within a complex sociolinguistic framework, wherein English simultaneously serves as the medium of higher education and a marker of their potential to partake in the global economy. At the same time, the Arabic language, Islamic values, and regional culture remain central to their personal and social selfhood. The association of AI-generated EFL content with Western social norms, androcentric professional roles, and settings that do not have cultural resonance is more than a mistake in representation. It can be considered as a marginalization of the identities of learners of Arabic, who do not see themselves in the materials through which they are invited to learn (Bender et al., 2021; Bencherrat & Baghor, 2025).

Before the ideological questions posed by RQ2 and RQ3 can be meaningfully answered, it is necessary to establish what linguistic features actually characterize AI-generated EFL reading texts at the level of surface form. This is the function of RQ1. Within Fairclough's (2023) model, claims about ideology and representation (Dimensions 2 and 3) must be grounded in a systematic account of textual choices (Dimension 1): pronoun distribution, transitivity patterns (who is positioned as agent versus patient), lexical collocation, and modality. Having analyzed these linguistic regularities throughout the corpus first, any further statements concerning gendered or cultural representation would be impressionistic rather than based on textual evidence. RQ1 is therefore not a peripheral or purely descriptive question, but rather it is the empirical basis upon which the ideological analysis that is the focus of RQ2 and RQ3 relies, and whose results (Section 4.1) are treated as load-bearing evidence in the Discussion rather than merely background description.

1.4 Purpose and Research Questions

In light of this, the current work employs Fairclough's (2023) three-dimensional CDA model to investigate the ideological representations for gender and Arab cultural identity in the EFL reading materials generated by three leading LLMs: ChatGPT-4o, Google Gemini 1.5 Pro, and Microsoft Copilot. In doing so, the research responds to a growing demand for culturally responsive evaluation frameworks for AI-generated learning content in non-Western higher education contexts. The following research questions will guide the study.

Linguistic features in RQ1 refer specifically to the four textual-level indicators operationalized in Phase A of the analysis (Section 3.4): gendered pronoun distribution, transitivity patterns (the grammatical positioning of male- and female-coded referents as agents or patients of action), lexical collocation (the adjectives, verbs, and nouns habitually paired with male or female referents), and cultural markers (names, places, institutional settings, and religious or cultural references). These features are included because they constitute the observable, codable evidence on which the ideological claims of RQ2 and RQ3 are built; without a systematic account of how gender and culture are linguistically encoded at the textual level, claims about representation and ideology would not be empirically traceable to the corpus. RQ1 is thus not an exploratory or subordinate question but rather a necessary and essential question. The quantitative and linguistic evidence provided by the findings of the study (in Section 4.1) support the qualitative CDA readings in Sections 4.2 – 4.6.

RQ1. What textual and linguistic features characterize AI-generated EFL reading materials when produced in response to neutral instructional prompts?

RQ2. How are gender and Arab/Islamic cultural identity represented discursively in AI-generated EFL reading texts?

RQ3. What ideological assumptions underpin the construction of knowledge, social roles, and identity in AI-generated EFL educational content?

The results are likely to have implications for essential discussions regarding AI literacy, equitable pedagogy, and the ethical use of AI tools in non-Western higher education settings, while also providing practical suggestions for EFL curriculum designers and institutional AI policy in the Gulf region.

2. Literature Review

2.1 Critical Discourse Analysis: Theoretical Foundations and Educational Applications

CDA is a very general interdisciplinary area of research that addresses the role of language in the social world as a resource. As an extension of the concept of language as a social practice, as put forward by Fairclough (2023), CDA sees language as a means of producing, reproducing, and resisting ideologies. He views language not only as a vehicle for communication, but a communicative activity as well. The model of three dimensions proposed by Fairclough is the dominant model used in the present study to capture the discourse. The model captures the discourse at three interrelated levels simultaneously. They are the textual dimension (lexis, grammar, transitivity, and modality), the discursive practice dimension (the text production, distribution, and consumption processes), and the sociocultural practice dimension (the institutional and ideological conditions embedding this discourse) of the text. This three-part framework is particularly useful in examining AI-generated educational materials, as it enables researchers to investigate how linguistic choices at the textual level are influenced by the AI's automated production at the discursive practice level, which in turn reveals societal power asymmetries at the sociocultural practice level.

Over recent years, educational research has begun to increasingly employ CDA in order to reveal how texts in pedagogic contexts reproduce dominant ideologies and shape learner identities. In a study by Simpson in 2025, a critical discourse analysis of AI guidance documents issued by six UK universities sheds light on the institutional language of efficiency, personalization, and student-centeredness. He argues that institutionally endorsed AI technologies become normalized under this institutional rhetoric while shielding human institutional actors from accountability. The argument that this discursive normalization of AI as a neutral administrative tool, as Simpson (2025) claims, is both insidious and dangerous. As Simpson (2025) argues, this discursive normalization of AI as a neutral administrative tool is insidious and dangerous. masks power imbalances that lie within automated decision-making systems and blocks critical assessment of the values and priorities that are programmed into them. In a recent study, Ting et al. (2025) utilized Fairclough's three-dimensional framework to examine the nature of AI-generated feedback on undergraduate EFL essays. They uncovered a six-part taxonomy of discourse moves, including diagnostic positioning, prescriptive directives, and affective engagement, demonstrating how AI technology exerts asymmetrical pedagogical authority over student writers. CDA studies indicate that human discourses, are of late, being interrogated to determine whether they can also tackle the AI-mediated textual environments, which are rapidly redefining education.

2.2 Generative AI in Language Education: Affordances and Pedagogical Concerns

There is a significant increase in scholarly interest in the use of generative tools such as ChatGPT, Google Gemini, and Microsoft Copilot in EFL and ESL. A lot of the literature examines their affordances for content generation, formative feedback, and learner scaffoldings. AI writing tools are "powerful scaffolding" for "textual and lexical development." However, they displace the writing struggle that is part of the process when learning a new language (Warschauer et al., 2023). In a systematic literature review, Bearman et al. (2023) observed a tension between the techno-optimist narratives of AI as democratizing technology in educational spaces and critical perspectives that emphasize the social, ethical, and pedagogical risks of delegating educational decisions to the automated.

The scrutiny of the educational political economy of AI is equally important. Knox's (2020) analysis of AI-in-education policy efforts in China reveals how national AI-education strategies reproduce technocratic governance ideologies while framing AI deployment as a politically-neutral efficiency tool that transcends everything education and pedagogy. Selwyn and Facer (2013) similarly argued that digital education technologies are not ideologically neutral instruments. Rather, they encode on the side of the designer's corporate or institutional professional, commercial, and political interests, and power relations. According to the scholars mentioned above, when EFL teachers rely on AI-produced reading materials in their classrooms, they are not adding neutral, context-free materials to the mix. Rather, they are importing a package of assumptions into the classroom about knowledge, social identity, and cultural normalcy that are encoded linguistically.

2.3 Algorithmic Bias and Ideological Representation in Large Language Models

A significant body of research has recently emerged that shows that large language models (LLMs) – the underlying technology of all major generative AI platforms – are not neutral content producers, but systems that encode and reproduce the ideological biases present in their training encyclopedias. According to a critical survey of bias in natural language processing by Blodgett et al. (2020), language technology is power. That is, the choices made by AI systems are never just technical decisions. They always engage in social hierarchies of race, gender, class, and culture. By analyzing different NLP systems, it was found that they consistently encode and amplify the biases of dominant-language majority users at the direct expense of marginalized populations. According to Bender et al. (2021), well known as the "stochastic parrots" paper that discusses the problems associated with the use of very large datasets for LLMs, such LLMs trained on vast internet crawls actually amplify the language and thinking of their most represented speakers. The latter were English-speaking,

Western, and male. In this way, these LLMs will not just under-represent the less powerful knowledge systems and languages but also misrepresent them. In addition, Dodge et al. (2021) provide direct empirical evidence for this claim: their systematic audit of the C4 corpus one of the world's largest datasets used to train contemporary LLMs found that 97.8 per cent of its documents are classified as White-aligned English. Furthermore, the corpus encodes measurable negative sentiment specifically associated with Arab identities making communities like those of Gulf EFL learners structurally invisible in the data from which AI educational content is produced.

The ideological dimensions of LLM outputs have been documented in educational discourse. Ahmed and Mahmood (2024) utilized Foucauldian discourse analysis along with Fairclough's CDA framework to analyze ChatGPT's responses with respect to disputable knowledge questions on Capitalism and Communism. The results suggest a clear bias towards a capitalist epistemology as the AI definition and framing normalized capitalist discourse as objective and making alternative discourses sound like ideologically suspect positions. According to Ahmed and Mahmood (2024), ChatGPT cannot be relied on for being biased in its responses (p. 184).

2.4 Gender, Culture, and Representation in AI-Generated Educational Texts

Research on gender, cultural representation and AI-generated language is significant but has remained under-researched till now. Kotek et al. (2023) conducted a study of gender bias and gender stereotypes in large language models (LLMs). It found that popular LLMs default to male pronouns and subject positions for agentic professional roles (like doctor, engineer, and executive) and female pronouns for nurturing, subordinate, or domestic roles. The result of these patterns is not coincidental, as it is a statistical reflection of the gender inequality encoded within the LLMs from their training data, which mediates generation of historical bias at scale in the content AI tools generate for classroom use. In an EFL educational context, such defaults are particularly important pedagogically. Indeed, if students are repeatedly shown AI-generated text that associates professional power with maleness, the language classroom can become a place where gender hierarchy is normalized.

The issue of cultural representation in AI-generated content is troublesome in non-Western education, too. Most AI-generated texts default to either Western social settings, Anglo-American naming conventions, secular value systems, or Euro-American institutional references (Bender et al 2021). For Arab EFL learners in Gulf universities who inhabit a quieter socio-cultural default shaped by Islamic values, the Arabic language, and GCC regional ones, this cultural default is not just a representational inconvenience. The phenomenon entails a continuous and systematic deterritorialization of their epistemic and cultural identities from the very materials through which they learn English. This study is the first to examine the gender and cultural representation in AI-generated EFL reading texts for Arab higher education settings. The literature is devoid of any study to empirically test such patterns so far.

3. Methodology

3.1 Research Design

The qualitative research design of this study is based on the Critical Discourse Analysis (CDA) tradition. Given that the study seeks to understand the ideological aspects, power relations, and representational patterns in AI-generated texts, such phenomena cannot be reduced to numbers. Qualitative inquiry helps me to engage with such a phenomenon (Creswell & Poth, 2018). To be more precise, the current research will adopt a document analysis strategy where the AI-generated EFL reading materials, which make up the purposefully constructed corpus, will function as the main data.

3.2 Theoretical Framework: Fairclough's Three-Dimensional Model

The current investigation of this research is based on Fairclough's three-dimensional model, which outlines CDA in a very illuminating and plausible manner. According to Fairclough, any discursive event can be conceptualized as simultaneously a text, a discursive practice, and a sociocultural practice.

Dimension 1-Text: The layer consists of careful linguistic analysis of AI-generated reading passages, including attention to choice of lexicon (word choice and connotations), transitivity (who acts, who is acted upon, in what processes), modality (levels of certainty and obligation encoded in auxiliary verbs), and pronouns (who is referred to, dis/empowerment through subject positions). For this dimension, technical linguistic tools have been provided by the System Functional Linguistics (SFL) of Halliday and Matthiessen (2014) which has aspects of transitivity and mood.

Dimension 2-Discursive Practice: how the texts are made, circulated, and used. The research centers on the circumstances surrounding LLM-generated content. This includes the impact of training data, algorithmic defaults, and the use of prompts and how they are framed. This production context affects the outcomes of the texts that teachers and learners are given and work with in the EFL classroom (Ting et al., 2025).

Dimension 3-Sociocultural Practice:

The discursive patterns observed in Dimensions 1 and 2 are situated in broader ideological and institutional contexts. They will show what the texts reveal about dominant assumptions regarding what is normal in the gender world, cultural normalcy, and the allocation of knowledge and agency across social groups (Ahmed & Mahmood, 2024).

3.3 Data Collection and Corpus Construction

The research study utilizes a purposively constructed corpus of AI-generated EFL reading materials, as per the researchers. Two experts estimate the quantified response of LLMs (ChatGPT 4.0, Gemini 1.5 Pro, Copilot) to the same standard prompts as equivalent. Each

prompt was submitted on each platform three times at different intervals of time to get different outputs, ultimately leading to the corpus of 45 texts (5 prompts $\times$ 3 platforms $\times$ 3 times). The following prompts were used:

1. "Write a 300-word EFL reading comprehension passage about a successful professional."
2. "Generate a reading text about family roles in modern society for B2 EFL learners."
3. "Create a reading passage about leadership and decision-making for university EFL students."
4. "Write an EFL reading text about science and discovery for Arab university students."
5. "Generate a discussion prompt about gender equality for an advanced EFL writing class."

The AI system could then see who they would choose to represent a particular role (male or female). It would give us an understanding of the choices that the AI system makes on its own. It will help us understand our AI system better. All outputs from AI were captured verbatim in real-time, time-stamped, and saved in a password-protected corpus file. The corpus file was distributed by platform, prompt, and iteration number. No editing was done to any of the output after its generation.

3.4 Analytical Procedure

The analysis was completed in three sequential but interrelated phases according to Fairclough (2023). As dictated by the established corpus-assisted CDA practice, "The full reproduction of exemplar texts enables full transparency and traceability of interpretation" (Baker et al., 2008).

Phase A. Textual Analysis: Each of the 45 texts was analyzed for four features: gendered pronoun distribution (e.g. did they use he/she or they and their possessive and object forms and how often); transitivity patterns (coding whether male-coded and female-coded referents ended up in grammatical subject agent or object patient positions); lexical collocations (which adjectives, verbs and nouns were most often associated with male or female referents); and cultural markers (i.e. names, places, institutional settings, cultural or religious references).

Phase B. Discursive Practice Analysis: This referred to the intertextual and interdiscursive features of the AI outputs, focusing on whose voice is discursively constructed as authoritative. How does AI position the implied learner-reader? And to what extent do the texts draw on Western default genres (for example, corporate professional narratives, nuclear family, secular scientific discourse)?

Phase C. Sociocultural Practice Analysis: The researchers analyzed the results from phase A and B against the sociocultural context of Gulf Arab higher education, leveraging the critical literature on algorithmic bias (Bender et al. 2021; Blodgett et al. 2020) and gendered AI (Kotek et al. 2023) to assess the ideological implications of the representational defaults uncovered in the corpus.

3.4.1 Quantification of Table 1 Variables

The percentages reported in Table 1 (Section 4.1) were generated as follows. For each of the ten variables listed in Table 1 (e.g., female protagonist in professional role, Western nuclear family as historical baseline, identical/near-identical texts across trials), the two trained coders independently applied a binary present/absent code to each relevant text, following the coding categories developed in Phase A and the codebook finalized after the reliability check described in Section 3.5. Variables tied to Prompt 1 (professional role) and Prompt 3 (leadership) were coded across the 3 trials generated by each of the 3 platforms ($n = 3$ texts per platform per prompt; $n = 9$ texts per prompt across all platforms). Variables tied to Prompt 2 (family roles) and Prompt 4 (scholarly heritage) were coded in the same manner within their respective 3-trial, 3-platform subsets ($n = 9$ texts each). The "Corpus Total" percentages combine the relevant nine-text subset for each variable, not all 45 texts, since each variable was meaningfully codable only within the texts generated by the prompt to which it pertains; this is why the corpus-level denominators in Table 1 are reported as $n = 9$ rather than $n = 45$. Raw counts (e.g., 2/3, 8/9) were converted to percentages by dividing the count by the total texts coded for that variable and multiplying by 100. Coding discrepancies between the two coders were resolved through the consensus process described in Section 3.5 prior to finalizing the counts reported in Table 1.

3.5 Validity and Reliability

To ensure the trustworthiness and credibility of the analysis, two independent coders who were both trained in CDA methodology and applied linguistics coded a random 20% subsample of the corpus (nine texts) using the coding framework developed for Phase A. According to Cofie et al. (2022), in qualitative CDA research, the measure of intercoder reliability should not be evaluated solely based on statistical measures of agreement, as is often typical, still it should rather take place via consensus processes based upon dialogue, which are epistemologically compatible with interpretivist inquiry. Discrepancies between the two coders were resolved through discussion, not overruled the second coder by the first coder, and a final codebook was created before coding of the rest of the corpus was completed. The eight-process-based guidelines for qualitative intercoder reliability by O'Connor and Joffe (2020) can help ensure consistency, reflexivity, communicability of the analytical framework among research team members do so. More specifically, researcher reflexivity was maintained throughout the study due to the analytical memo journal in which both coders documented interpretive decisions and possible sites of bias at every phase of the analysis.

4. Analysis and Discussion

The findings in this section are presented in two layers. To begin with, the overall provision of pronouns, identities in protagonists, and

cultural markers is given for the entire 45-text corpus. Then it is followed by an analysis of three-dimensional CDA of three exemplar texts chosen for their analytically productive context, followed by a composite-dimensional discussion that integrates all findings.

4.1 Quantitative Overview of the Corpus

The study subjected all 45 texts to systematic analysis across five prompts, three platforms, and three trials per platform, revealing patterns that confirmed but also complicated the initial hypotheses of the study. Table 1 provides the main aggregated findings by prompt category and platform.

Table 1. Summary of Protagonist Identity, Pronoun Distribution, and Cultural Marker Frequencies Across the 45-Text Corpus

Variable	ChatGPT-4o (n=15)	Gemini 1.5 Pro (n=15)	Microsoft (n=15)	Copilot	Corpus Total
Female protagonist in professional role (P1)	2/3 (67%)	3/3 (100%)	3/3 (100%)		8/9 (89%)
Male protagonist in professional role (P1)	1/3 (33%)	0/3 (0%)	0/3 (0%)		1/9 (11%)
Arab/regional names in professional texts (P1)	3/3 (100%)	1/3 (33%)	0/3 (0%)		4/9 (44%)
Western names as default in professional texts (P1)	0/3 (0%)	2/3 (67%)	3/3 (100%)		5/9 (56%)
Named protagonist in leadership texts (P3)	0/3 (0%)	0/3 (0%)	0/3 (0%)		0/9 (0%)
Western nuclear family as historical baseline (P2)	3/3 (100%)	3/3 (100%)	3/3 (100%)		9/9 (100%)
Arab/Islamic family structures referenced (P2)	0/3 (0%)	0/3 (0%)	0/3 (0%)		0/9 (0%)
Arab scholarly heritage specifically named (P4)	2/3 (67%)	3/3 (100%)	0/3 (0%)		5/9 (56%)
Identical/near-identical texts across trials (P1/P3)	0/3 (0%)	1/3 (33%)	3/3 (100%)		4/9 (44%)
Gender equality framed within Western feminist discourse (P5)	3/3 (100%)	3/3 (100%)	3/3 (100%)		9/9 (100%)

Four major findings can be obtained from the data aggregate. Initially, contrary to our expectations, all three platforms assigned female protagonists to the neutral “successful professional” prompt (89 %), with ChatGPT naming a male in one trial, the only gendered case across systems. No platform either assigned a named or gendered protagonist to any of the nine leadership texts: the leadership corpus is uniformly abstract and impersonal. In a universal invocation of historical living arrangements, the Western patriarchal nuclear family is nevertheless the only baseline. The Western liberal feminist discourse is used in framing the fourth prompt and all the other prompts. The subsequent detailed investigations are based on these four patterns.

4.2 Close Analysis of Text 1 - ChatGPT-4o, GPT-P1-T1: The Arab Female Professional and the Relational Achievement Script

Text 1 (reproduced, GPT-P1-T1):

Aisha Al-Harthy is a well-known engineer in Oman who has achieved remarkable success in her field. From a young age, Aisha showed a strong interest in mathematics and problem-solving. She studied hard in school and earned a scholarship to study engineering at a top university. Although the program was challenging, she remained focused and determined. On graduation, Aisha went on to work for an international construction firm. Initially she struggled with having to work in a high-speed environment. She soon learned new skills and developed her confidence though. She excelled in her role due to her capacity to handle intricate projects and successfully collaborate with her coworkers. The most significant success Aisha has had is leading a project to create a sustainable energy plant. The project was designed to improve the quality of the environment and to provide clean energy for use in the local community. Her leadership and creative thinking enabled the project to be a success. Today, Aisha is respected not only for her professional accomplishments but also for her commitment to helping others. She frequently mentors young students, specifically women, on a regular basis to help encourage them to go into a career in science and engineering. Aisha feels that it is important to work hard, never give up, and learn from failure to succeed.

Textual Analysis: Aisha Al-Harthy is an Arab female protagonist, and the narrative is clearly set in Oman – the same country that is home to the protagonist – which is a deliberate break with the Western male default we might anticipate, and a clear reference to gender and cultural inclusion. Although real, Aisha’s professional agency is discursively bounded in an achievement script of relational doing, a

closer transitivity analysis suggests. The material process clauses that go with that Aisha are she studied, she earned, she joined, she learnt, she gained confidence, she manages all agentive and positive. The result of the discursive representation of her success is not the attainment of professionalism but service and mentorship to the community, as she regularly mentors young students, especially women. At the closing textual position, i.e., the discursively most prominent clause, her greatest representational value is as a guide for other women. The same adjective cluster collocating with Aisha codes for success by virtue of being focused, determined, and committed to helping others. This pattern consistently repeats itself in the cases of both female protagonists generated by ChatGPT. Sara Al-Harthy (GPT-P1-T2) “GPT-P1-T2 closes identically: ‘Sara Al-Harthy... often speaks at conferences and encourages young people, especially women, to pursue careers in STEM. She believes success is built on community, not competition.’ The discursive closure female success validated through communal reproduction is literally identical to GPT-P1-T1.”

Contrastive Transitivity - GPT-P1-T3 (Ahmed Al-Harthy). The single male protagonist text generated by ChatGPT sees Ahmed Al-Harthy occupying the subject position of a drastically different success narrative. The focus in Ahmed's story is placed on the individual, the struggles, and difficulties he faced, and the perseverance on his part: “*He was turned down a lot of times, and this made him very sad, but he did not give up. He eventually got a job.*” Ahmed does not have the same involvement in mentoring and encouraging women as Aisha and Sara do. His success turns into a triumph over adversity for him. Male and female success are articulated in a well-established binary in professional discourse: Female success is a relational contribution, whereas male success is individual perseverance (Kotek et al 2023).

Discursive Practice Analysis: At the level of discursive practice, the second aspect is the frequency of the phrase especially women, and the relational success framework in two independent trials of ChatGPT, indicating a stylistic feature that is statistically produced. The training data clearly contains a schema of the ‘successful woman in STEM,’ which is reflected in the kinds of women-in-science profiles it is likely to find in Western media and will almost certainly reproduce in the prompts to which it is presented.

Sociocultural Practice Analysis: The text at the sociocultural level is an ideologization of a complicated nature. On the contrary, Aisha Al-Harthy, sustainable energy, local communities, Oman's cultural specificity are components that lend a degree of regional resemblance to most AI-generated EFL that lacks. The professional success narrative via which Aisha is constructed, scholarship → international company → community mentor. On the other hand, follow the template of Western meritocratic individualism with no recognition of the specific institutional, family, or cultural conditions which shape women's professional trajectories in the Gulf. Even if the sustainability theme is locally relevant, the green-technology discourse embedded in the strategy has a universal, rather than an Omani development character, reinforcing Selwyn and Facer's (2013) argument that AI tools impart the values and priorities of their dominant user communities, regardless of surface-sociolinguistic customization.

4.3 Close Analysis of Text 2 - Gemini 1.5 Pro, GEM-P1-T1: The Western Female Default and Algorithmic Intertextuality

Text 2 (reproduced, GEM-P1-T1, excerpt):

Elena Rossi did not always want to be an architect. Growing up in a small coastal town in Italy, she spent her days sketching the way sunlight hit the ancient stone walls of her village. Today, she is one of the world's leading figures in sustainable urban design... For Elena, success is not measured by the height of a skyscraper, but by its carbon footprint... She serves as a powerful example that with a clear vision and challenging work, one person can profoundly change the landscape of the future.

Textual and Intertextual Analysis. Gemini's professional texts stand out analytically in two dimensions. The protagonist's name is Elena Rossi. She is Italian and European, and the scene is set in Milan, London, and other European cities and locations. The selection of an Italian architect within a cosmopolitan European context suggests that Western Europe's professional world is the universal model for professional success. Secondly, and crucially, the Elena Rossi character appears verbatim in two of Gemini's three trials (GEM-P1-T1 and GEM-P3-T3), with the same backstory “growing up in a small coastal town.” The intertextual repetition across different sessions, same character uncovers an extremely high degree of algorithmic anchoring at play: Gemini appears to have a strongly activated “successful female professional” template that resurfaces consistently.

Such a conclusion has great implications for EFL classrooms, whereby a constant exposure to the same Western female archetype as the role model professionalizes the European identity as the ubiquitous face of success. The discursive construction of the success of Elena, further, is reminiscent of the relational achievement script that was found in GPT-P1-T1: “she still finds time to mentor young students, and her success is defined in terms of the highest value when it benefits the greater good of the planet.”

4.4 Close Analysis of Text 3 - The Leadership Corpus: Gender Erasure as Ideological Practice

Text 3 (representative excerpt, GPT-P1-T1, Prompt 3):

Effective leadership is intricately linked to the ability to make sound decisions. In academic and professional settings, leaders are often required to evaluate complex situations and choose the best course of action. Decision Making is a process of obtaining information, evaluating options, and weighing the outcomes. A good leader does not just depend on instinct; it is a mix of experience and thoughtfulness. They also know the importance of seeking input from others and respecting other viewpoints.

The most striking corpus-level observation in the leadership category is that there is no named or gendered main character in any of the nine texts on leadership. Each medium and each experiment generated abstract, impersonal rhetoric on leadership whereby the “leader” is called upon as a generic, a-historical name. In contrast to the professional success texts, where the AI systems have come up with named persons, leadership is only achieved by nominalizations (decision-making, leadership, communication), generic noun phrases (strong

leaders, effective leaders, skilled leaders), and inclusive plural pronouns (they, their). This discursive de-personalization of leadership speech is nothing linguistically neutral; it is a discursive specific option, whose ideological implications can be outlined. The texts eliminate the historical and empirical fact that leadership roles are highly gendered, racialized, and culturally stratified spaces by rendering leadership a set of disembodied, universally applied competencies, homogenized (Kotek et al., 2023; Blodgett et al., 2020). Arab EFL learners, who live in institutions in which leadership succession, authority, and governance are mediated by certain cultural, religious, and family organization, are exposed to leadership in these texts as a body-free, culturally decontextualized set of skills. The tacit message is that the “ideal leader” is located nowhere culturally and ironically restates the universalizing presuppositions of the Western view of organizational theory as the default frame.

4.5 Family Roles and the Western Developmental Trajectory

In each of the three texts on all nine family roles, the invocation of “the traditional family,” i.e., the breadwinning father and the housewife, is the historical universal on which the “modern” families have, positively, developed. This discursive form plays the role that Fairclough (2023) would define as a naturalization strategy: by making the Western nuclear family of Patriarchal society the historical standard, the texts dictate the massive family diversity of extended kinship, Islamic jurisprudential systems of domestic responsibility (qiwamah), and tribal/communal-based networks of childcare that is the reality of Arab and Gulf family life. No literature in the corpus recognizes the possibility that the assumption that family roles are altering does not apply or should not apply to all cultures.

4.6 Cross-Dimensional Discussion: Ideology, Overcorrection, and Epistemic Exclusion

Combined, the results of the actual corpus demonstrate a more convoluted and, in several ways, more pernicious ideological construction than a mere male- default or Western- default tale. There are three analytical conclusions.

First, the “overcorrection” toward female professionals, though seemingly progressive on its face, reinstates gender ideology via unequal discursive construction: female success is always constructed as relational, community-oriented, and mentorship-oriented, and the single male professional narrative as that of individual persistence and resilience. This achievement discourse binary of male/female is not only ingrained in the training data of the LLMs (Bender et al., 2021) but also continues to be present even when the AI supposedly does not default to the most blatantly male way.

Second, the total elimination of named, embodied protagonists in discourses about leadership is a particular kind of gender and cultural erasure: through the complete abstraction of leadership, all three platforms generate texts that do not give Arab EFL learners any usable model of culturally embodied leadership, no Arab leader, no female leader, no Gulf institutional context. The authoritative and impersonal strong leaders of these texts are those who talk out of no one and nowhere and in the process, take a Western universalist stance that is masqueraded as neutrality (Hofstede, 2001).

Third, the stability of the Western nuclear family developmental discourse through 100% of the family roles texts, plus the total absence of the Islamic or Arab family formations, conducts a systematic erasure of epistemic representations of the domestic world of Arab learners. These students are encouraged to consider their own family experiences as “traditional” (and implicitly backward) when compared to a Western liberal modernity as the universal goal of social progress (Selwyn & Facer, 2013). These results have direct implications for EFL curriculum design, teacher education, and institutional AI policy in the situations of the Gulf higher education, which are discussed in the subsequent section.

5. Conclusion and Implications

5.1 Revisiting the Research Questions

In terms of RQ1, AI-generated EFL still does not make linguistically neutral texts: the female professionals are built with relational lexis, texts about leadership are depersonalized, and texts about family roles use the Western nuclear family as the historical reference point. In relation to RQ2, there is apparent cultural inclusiveness and underlying representational inequalities: the success of women is collective and relational, whereas the success of men is individual and resilient; the Arab cultural identity is a historical background and not a contemporary living location. In relation to RQ3, the corpus is supported by three ideological formations: a Western liberal feminist ideology that maintains gendered difference in discussions of achievements; a Western developmental framework that makes Arab structures of the family invisible; and a universalist organizational ideology that removes cultural and gendered specificity from leadership.

5.2 Implications

Pedagogically, teachers of EFL in Gulf universities will have to acquire critical AI literacy - the ability to assess AI-generated texts as ideological representation, rather than linguistic accuracy. The pre-deployment screening protocol with the service of gendered transitivity, cultural indicators, and the construction of controversial social concepts would help to mitigate the risks of unthinkingly blindly transferring ideologically colored materials into the classroom (Ting et al., 2025; Ahmed & Mahmood, 2024).

Technologically, context-sensitive localization mechanisms in the field of education should be created to permit institutional users to define cultural and demographic parameters for AI platforms. The discovery of algorithmic cloning, specifically the almost verbatim reproduction of the identical texts by Copilot across multiple trials, should be technologically investigated, since it is necessary to have output variance in educational use cases (Holtzman et al., 2020).

At the policy level, Gulf higher education institutions should require a critical discourse review of AI-generated instructional content before classroom deployment. Steps should be taken at the national level to create region-specific AI content standards that reflect the cultural and linguistic contexts of Arab EFL learners (Bearman et al., 2023; Selwyn & Facer, 2013).

5.3 Limitations and Future Directions

This corpus instantiates AI output at one point in time. Representation patterns may shift as models are updated. The investigation pertains exclusively to English prompts, leaving the comparison with Arabic prompts as a productive future direction. It is also important to point out that learner reception studies and longitudinal tracking of platform-specific representational change are the most critical avenues for future studies.

Acknowledgments

Not Applicable.

Author contributions

Dr. Adnan Mohammed Bataineh, Mr. Kashif Ali Sabiri, and Mr. Ebrahim Douzandeh contributed equally to the conception and design of the study, literature review, data collection, data analysis and interpretation, and preparation and critical revision of the manuscript. All authors read, approved the final manuscript, and agreed to be accountable for all aspects of the work. The authors confirm that they contributed equally to the study.

Funding

This research was supported by the Deanship of Postgraduate Studies, Research, and Innovation, University of Buraimi under Grant Number UoB\RIEU\IRG\2024-25\CFS\001. The funder had no role in the study design, data collection, analysis, or the decision to submit the work for publication.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The authors confirm that this research abides by the ethical standards for research involving human subjects. Participation was voluntary, and anonymity and confidentiality of the participants was strictly respected throughout the research process. The study had minimal risk, and no sensitive personal data was involved and thus was carried out in accordance with the general ethical guidelines of the authors' host institution.

Provenance and peer review

Not commissioned, externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data is not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data is available.

Open access

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

Copyrights

Copyright for this article is retained by the authors, with first publication rights granted to the journal.

AI Usage Disclosure

The authors utilized Claude (free version) for preparing this manuscript, only for language polishing (grammar, spelling, and punctuation), of the Introduction and Literature Review. The tool did not produce or select references, nor did it produce or interpret Methodology, Analysis and Discussion or the Conclusion sections of the work, nor did it produce research hypotheses or original analysis. All of the suggestions made with the help of AI were critically reviewed and edited by the authors who are entirely responsible for the accuracy, integrity, and originality of the manuscript.

References

Ahmed, T. N., & Mahmood, K. A. (2024). A critical discourse analysis of ChatGPT's role in knowledge and power production. *Arab*

World English Journal, Special Issue on ChatGPT, 184-196. <https://doi.org/10.24093/awej/ChatGPT.12>

- Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyżanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, 19(3), 273-306. <https://doi.org/10.1177/0957926508088962>
- Bearman, M., Ryan, J., & Ajjaw, R. (2023). Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education*, 86(2), 369-385. <https://doi.org/10.1007/s10734-022-00937-2>
- Bencherrat, N., & Baghor, Y. (2025). Language and artificial intelligence in Arab societies: A sociolinguistic approach to identity and digital hegemony. *Journal of Studies in Language, Culture, and Society*, 8(4), 279-297. <https://asjp.cerist.dz/en/article/286512>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>
- Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (technology) is power: A critical survey of “bias” in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5454-5476). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.485>
- Cofie, N., Braund, H., & Dalgarno, N. (2022). Eight ways to get a grip on intercoder reliability using qualitative-based measures. *Canadian Medical Education Journal*, 13(2), 73-76. <https://doi.org/10.36834/cmej.72504>
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
- Daryani, Y., Sourati, Z., & Dehghani, M. (2026). The homogenizing engine: AI’s role in standardizing culture and the path to policy. *Policy Insights from the Behavioral and Brain Sciences*, 13(1), 14-27. <https://doi.org/10.1177/23727322251406591>
- Dodge, J., Sap, M., Marasović, A., Agnew, W., Ilharco, G., Groeneveld, D., Mitchell, M., & Gardner, M. (2021). Documenting large webtext corpora: A case study on the Colossal Clean Crawled Corpus. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 1286-1305). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.emnlp-main.98>
- Fairclough, N. (2023). Critical discourse analysis. In J. P. Gee and M. Handford (Eds.), *The Routledge handbook of discourse analysis* (2nd ed., pp. 11-22). Routledge. <https://doi.org/10.4324/9781003035244-3>
- Fenech-Borg, E. Z., Meznaric-Kos, T. P., Lekovic-Bojovic, M. D., & Hentze-Djurhuus, A. J. (2025). *The cultural gene of large language models: A study on the impact of cross-corpus training on model values and biases* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2508.12411>
- Ge, Z.-G., Liu, Y., & Zhao, G. (2025). Exploring the utilization of AI-generated learning resources: A systematic review. *AI-Enhanced Learning*, 1(2), 295-325. <https://doi.org/10.70725/060035hqdeea>
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday’s introduction to functional grammar* (4th ed.). Routledge.
- Hofstede, G. (2001). *Culture’s consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). SAGE Publications.
- Holtzman, A., Buys, J., Du, L., Forbes, M., & Choi, Y. (2020). The curious case of neural text degeneration. In *International Conference on Learning Representations*. OpenReview. <https://openreview.net/forum?id=rygGQyrFvH>
- Huang, Q., Lv, C., Lu, L., & Tu, S. (2025). Evaluating the quality of AI-generated digital educational resources for university teaching and learning. *Systems*, 13(3), Article 174. <https://doi.org/10.3390/systems13030174>
- Knox, J. (2020). Artificial intelligence and education in China. *Learning, Media and Technology*, 45(3), 298-311. <https://doi.org/10.1080/17439884.2020.1754236>
- Kotek, H., Dockum, R., & Sun, D. Q. (2023). Gender bias and stereotypes in large language models. In *Proceedings of the ACM Collective Intelligence Conference* (pp. 12-24). Association for Computing Machinery. <https://doi.org/10.1145/3582269.3615599>
- Mushtaq, A., Taj, I., Naeem, R., Ghaznavi, I., & Qadir, J. (2026). WorldView-Bench: A benchmark for evaluating global cultural perspectives in large language models. *Journal of Artificial Intelligence Research*, 85, Article 45. <https://doi.org/10.1613/jair.1.19001>
- O’Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, 19, 1-13. <https://doi.org/10.1177/1609406919899220>
- Selwyn, N., & Facer, K. (Eds.). (2013). *The politics of education and technology: Conflicts, controversies, and connections*. Palgrave Macmillan. <https://doi.org/10.1057/9781137031983>
- Simpson, N. H. (2025). Framing AI in higher education: A critical discourse analysis of inclusivity and power in university AI guidance documents. *Educational Linguistics*, 4(1), 33-53. <https://doi.org/10.1515/eduling-2024-0007>

Ting, L., Singh, C. K. S., Kiong, T. T., Rakhimjonov, F., & Singh, T. S. M. (2025). Mentor or examiner? A critical discourse analysis of AI-generated feedback in EFL writing education. *International Journal of Academic Research in Progressive Education and Development*, 14(4), 664-684. <https://doi.org/10.6007/IJARPED/v14-i4/26651>

Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. *Journal of Second Language Writing*, 62, Article 101071. <https://doi.org/10.1016/j.jslw.2023.101071>

Appendix A

The Prompts Used

1. “Write a 300-word EFL reading comprehension passage about a successful professional.”
2. “Generate a reading text about family roles in modern society for B2 EFL learners.”
3. “Create a reading passage about leadership and decision-making for university EFL students.”
4. “Write an EFL reading text about science and discovery for Arab university students.”
5. “Generate a discussion prompt about gender equality for an advanced EFL writing class.”

Appendix B

Table 1. Summary of Protagonist Identity, Pronoun Distribution, and Cultural Marker Frequencies Across the 45-Text Corpus

Variable	ChatGPT-4o (n=15)	Gemini 1.5 Pro (n=15)	Microsoft (n=15)	Copilot	Corpus Total
Female protagonist in professional role (P1)	2/3 (67%)	3/3 (100%)	3/3 (100%)		8/9 (89%)
Male protagonist in professional role (P1)	1/3 (33%)	0/3 (0%)	0/3 (0%)		1/9 (11%)
Arab/regional names in professional texts (P1)	3/3 (100%)	1/3 (33%)	0/3 (0%)		4/9 (44%)
Western names as default in professional texts (P1)	0/3 (0%)	2/3 (67%)	3/3 (100%)		5/9 (56%)
Named protagonist in leadership texts (P3)	0/3 (0%)	0/3 (0%)	0/3 (0%)		0/9 (0%)
Western nuclear family as historical baseline (P2)	3/3 (100%)	3/3 (100%)	3/3 (100%)		9/9 (100%)
Arab/Islamic family structures referenced (P2)	0/3 (0%)	0/3 (0%)	0/3 (0%)		0/9 (0%)
Arab scholarly heritage specifically named (P4)	2/3 (67%)	3/3 (100%)	0/3 (0%)		5/9 (56%)
Identical/near-identical texts across trials (P1/P3)	0/3 (0%)	1/3 (33%)	3/3 (100%)		4/9 (44%)
Gender equality framed within Western feminist discourse (P5)	3/3 (100%)	3/3 (100%)	3/3 (100%)		9/9 (100%)