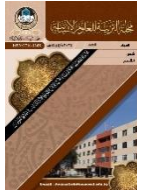




Journal of Education for Humanities

A peer-reviewed quarterly scientific journal issued by College of Education for Humanities / University of Mosul



Bridging Minds and Machines: The Role of English in Advancing AI Technology

Narjis Audah Rashk ¹

Estabraq R. Ibrahim ²

Misan University / College of Basic Education /English Department ¹

Baghdad College of Economic Sciences University, English department ²

Article information

Accepted : 27/11/2025

Published 19/6/2026

Keywords:

English as a Lingua Franca, Artificial Intelligence (AI), Large Language Models (LLMs), Low-Resource Languages

Correspondence:

Narjis Audah

Narcissusodaa@uomisan.edu.iq

Abstract

The rapid integration of Artificial Intelligence (AI) into diverse fields has reinforced the critical role of English as both a global lingua franca and a medium for technological advancement. This study, *Bridging Minds and Machines: The Role of English in Advancing AI Technology*, investigates how English functions as the dominant linguistic framework for AI research, programming, and human-computer interaction. The research highlights the centrality of English in major AI hubs such as OpenAI (San Francisco, USA), DeepMind (London, UK), and Google AI (Mountain View, USA), where English underpins the design of large language models (LLMs), including GPT-5 and Gemini. It further examines how English-based corpora, such as the Common Crawl and Wikipedia, form the backbone of AI training datasets, while acknowledging the challenges of linguistic bias, marginalization of low-resource languages, and the need for more inclusive multilingual models. By drawing on case studies from Iraq and the broader Middle East, the study contextualizes how English proficiency among researchers and students influences access to AI scholarship, coding communities (e.g., GitHub, Kaggle), and open-source collaborations. Additionally, it explores pedagogical

implications, particularly the integration of English for Specific Purposes (ESP) in university curricula to prepare future generations of Iraqi scholars and developers to engage with global AI discourses. The findings argue that while English continues to be the dominant bridge between human minds and intelligent machines, a critical approach is required to balance this dominance with the promotion of linguistic diversity. This study contributes to both applied linguistics and digital studies by showing how English is not only a communication tool but also an epistemic gatekeeper in the age of AI

DOI: *****, ©Authors, 2025, College of Education for Humanities University of Mosul.

This is an open access article under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).



مجلة التربية للعلوم الإنسانية

مجلة علمية فصلية محكمة، تصدر عن كلية التربية للعلوم الإنسانية / جامعة الموصل



الربط العقول والآلات: دور اللغة الإنجليزية في تطوير تكنولوجيا الذكاء الاصطناعي

نرجس عودة رشك¹ ، استبرق ر. إبراهيم²

جامعة ميسان / كلية التربية الأساسية / قسم اللغة الإنجليزية¹

كلية بغداد للعلوم الاقتصادية الجامعة / قسم اللغة الإنجليزية²

الملخص

معلومات الارشفة

أدى الاندماج السريع للذكاء الاصطناعي في مجالات متعددة إلى تعزيز الدور الحيوي للغة الإنجليزية بوصفها لغة عالمية مشتركة وسيلة للتقدم التكنولوجي في الوقت نفسه. وتبحث هذه الدراسة، الموسومة بـ الربط بين العقول والآلات: دور اللغة الإنجليزية في تطوير تكنولوجيا الذكاء الاصطناعي، في كيفية عمل اللغة الإنجليزية بوصفها الإطار اللغوي المهيمن في أبحاث الذكاء الاصطناعي، والبرمجة، والتفاعل بين الإنسان والحاسوب.

تاريخ القبول : 2025/11/27

تاريخ النشر : /1962026/

الكلمات المفتاحية :

اللغة الإنجليزية بوصفها لغة مشتركة، الذكاء الاصطناعي، النماذج اللغوية الكبيرة، اللغات محدودة الموارد.

وتسلط الدراسة الضوء على مركزية اللغة الإنجليزية في مراكز الذكاء الاصطناعي الكبرى، مثل OpenAI في سان فرانسيسكو بالولايات المتحدة الأمريكية، و DeepMind في لندن بالمملكة المتحدة، و Google AI في ماونتن فيو بالولايات المتحدة الأمريكية، إذ تشكل اللغة الإنجليزية أساسًا في تصميم النماذج اللغوية الكبيرة، ومنها GPT-5 و Gemini. كما تبحث الدراسة في كيفية تشكيل المدونات النصية القائمة على اللغة الإنجليزية، مثل Common Crawl و Wikipedia، العمود الفقري لمجموعات البيانات المستخدمة في تدريب الذكاء الاصطناعي، مع الإقرار في الوقت نفسه بالتحديات المرتبطة بالتحيز اللغوي، وتهميش اللغات محدودة الموارد، والحاجة إلى نماذج متعددة اللغات أكثر شمولًا. ومن خلال الاستناد إلى دراسات حالة من العراق والشرق الأوسط عمومًا، تضع الدراسة في سياقها أثر إتقان اللغة الإنجليزية لدى الباحثين والطلبة في الوصول إلى أدبيات الذكاء الاصطناعي، ومجتمعات البرمجة، مثل GitHub و Kaggle،

معلومات الاتصال

نرجس عودة رشك

Narcissusodaa@uomisn.edu.iq

والتعاونات مفتوحة المصدر. كما تستكشف الدراسة الآثار التربوية لهذا الموضوع، ولا سيما دمج اللغة الإنجليزية لأغراض خاصة في المناهج الجامعية، بما يسهم في إعداد الأجيال القادمة من الباحثين والمطورين العراقيين للانخراط في الخطابات العالمية المتعلقة بالذكاء الاصطناعي. وتلخص الدراسة إلى أن اللغة الإنجليزية، على الرغم من استمرارها بوصفها الجسر المهيمن بين العقول البشرية والآلات الذكية، تحتاج إلى مقارنة نقدية توازن بين هذا الحضور المهيمن وبين تعزيز التنوع اللغوي. وتسهم هذه الدراسة في حقل اللسانيات التطبيقية والدراسات الرقمية من خلال إظهار أن اللغة الإنجليزية ليست مجرد أداة تواصل، بل تمثل أيضًا بوابة معرفية في عصر الذكاء الاصطناعي.

DOI: *****, ©Authors, 2025, College of Education for Humanities University of Mosul.

This is an open access article under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

1.1 Background of the Study

The twenty-first century has been marked by unprecedented technological advancements, with Artificial Intelligence (AI) emerging as a transformative force across multiple disciplines, from healthcare and education to finance and communication. Central to this transformation is the role of language, which functions as the primary medium of interaction between humans and machines. Among global languages, English has become the dominant linguistic code in AI development, shaping how algorithms are trained, how users interact with AI systems, and how research is disseminated.

Leading AI research institutions, such as OpenAI in San Francisco, DeepMind in London, and Google AI in Mountain View, predominantly rely on English-based resources to design and refine advanced models like GPT-5, Gemini, and Claude. Training datasets such as the Common Crawl, Wikipedia, and BooksCorpus reflect a heavy reliance on English, establishing it as the epistemic foundation of modern

AI. While this dominance accelerates innovation and provides a unified framework for global collaboration, it also raises issues of linguistic bias, underrepresentation of low-resource languages, and unequal access to AI knowledge. In regions like the Middle East, and particularly Iraq, the implications of this dynamic are profound. English proficiency becomes not merely a skill but an entry point to global coding communities (e.g., GitHub, Kaggle), AI research networks, and academic scholarship. For Iraqi researchers and students, the mastery of English is often synonymous with the ability to participate in cutting-edge AI development and discourse.

The dominance of English thus functions as both a bridge and a barrier facilitating global communication while simultaneously excluding those without adequate linguistic preparation. Moreover, the integration of English in AI development extends beyond technical documentation and programming languages; it shapes the very structure of knowledge creation and dissemination. Conferences, webinars, and scholarly journals in AI overwhelmingly employ English as the lingua franca, creating an environment in which non-native speakers must navigate not only linguistic challenges but also cultural nuances embedded in academic communication.

For Iraqi academics and practitioners, this necessitates the acquisition of specialized English skills, ranging from technical writing and scientific argumentation to understanding idiomatic expressions and conventions of scholarly discourse. Consequently, proficiency in English becomes a strategic tool, enabling access to research collaborations, grant opportunities, and participation in global AI forums, thereby positioning English as a gatekeeper to knowledge and innovation in the twenty-first century. Simultaneously, the reliance on English-centric AI frameworks highlights broader socio-technological inequalities. Low-

resource languages, including Arabic dialects spoken across Iraq, are underrepresented in datasets, resulting in AI systems that may struggle with local linguistic and cultural contexts. This underrepresentation can affect the development of AI applications tailored to Iraqi society, from automated translation and voice recognition to educational tools and healthcare solutions. As a result, English dominance in AI does not merely reflect linguistic preference but has tangible consequences for technological inclusion and equity. Addressing these disparities requires deliberate policy interventions, localized data collection, and capacity-building programs in Iraq to ensure that

AI technologies serve diverse linguistic communities and do not reinforce global knowledge hierarchies.

1.2 Statement of the Problem

Although AI technologies are rapidly advancing, the linguistic dominance of English creates a disproportionate imbalance in participation and innovation. Most large-scale AI systems are designed around English-based linguistic structures, resulting in underrepresentation of Arabic and other regional languages in training corpora. This imbalance produces practical challenges: AI tools in Iraq and the Arab world often perform poorly in processing Arabic dialects, misinterpret cultural nuances, and fail to represent local knowledge. Moreover, while English serves as the “language of access” to AI, many Iraqi students and professionals face barriers due to limited English proficiency, outdated curricula, and a lack of localized AI resources. Consequently, the role of English in advancing AI technology is paradoxical: it enables global collaboration and innovation while simultaneously restricting equitable access for non-native speakers.

To address this paradox, there is a growing need for localized AI education and resource development in Iraq. Initiatives that integrate Arabic-language datasets, culturally relevant examples, and bilingual instructional materials can empower students and professionals to engage meaningfully with AI technologies without being hindered by linguistic barriers. Universities, research centers, and tech hubs in Iraq can play a pivotal role by offering specialized courses in AI literacy, computational linguistics, and English for AI purposes, thereby bridging the gap between global innovation and local participation. By fostering both English proficiency and native-language support, Iraq can cultivate a generation of AI practitioners who are capable of contributing to international research while developing tools and applications tailored to the linguistic and cultural context of their own society.

1.3 Research Objectives

This study aims to:

1. Examine the role of English as the dominant language in AI research, programming, and human-machine interaction.
2. Analyze the implications of English dominance on linguistic inclusivity, particularly in relation to low-resource languages such as Arabic.
3. Explore how English proficiency influences Iraqi researchers’ and students’ participation in AI-related fields.
4. Suggest strategies for balancing the dominance of English with the promotion of linguistic diversity in AI technologies.

1.4 Research Questions

1. How does English function as a linguistic framework in the advancement of AI technologies globally?
2. What challenges arise from the reliance on English-based corpora for AI training, particularly for low-resource languages such as Arabic?
3. In what ways does English proficiency affect Iraqi scholars' and students' ability to engage with AI research and applications?
4. What strategies can be adopted to promote multilingual inclusivity in AI development while maintaining English as a global bridge?

1.5 Significance of the Study

This study is significant for several reasons. Academically, it contributes to applied linguistics and digital studies by highlighting how language not only facilitates communication but also acts as a gatekeeper of knowledge and innovation. Practically, it provides insights for policymakers, educators, and technology developers in Iraq and the Middle East who seek to integrate AI technologies into education, research, and industry. For universities, the findings underscore the need to integrate English for Specific Purposes (ESP) courses tailored to AI and computer science students, ensuring they can actively participate in global research. For developers, the study emphasizes the necessity of multilingual AI systems that reflect cultural and linguistic diversity. On a broader scale, this research highlights the urgency of democratizing AI development so that technological progress does not reinforce existing global inequalities.

2: Theoretical Background and Literature Review

2.1 Artificial Intelligence

Artificial Intelligence (AI) refers to the branch of computer science that aims to create systems capable of performing tasks that traditionally require human intelligence, such as reasoning, learning, problem-solving, and language understanding (Russell & Norvig, 2021). Over the past few decades, AI has evolved from simple rule-based programs to sophisticated machine learning and deep learning models that can analyze large datasets, recognize patterns, and make predictions with increasing accuracy.

This evolution has transformed multiple sectors, including healthcare, finance, education, and communication, by automating processes, enhancing decision-making, and generating insights that were previously unattainable (Tegmark, 2017). A key aspect of modern AI is its reliance on data-driven learning. Systems such as large language models, including GPT-5, Claude, and Gemini, are trained on extensive datasets like Common Crawl, Wikipedia, BooksCorpus that.

enable them to understand and generate human-like text (OpenAI, 2024). This capability allows AI to perform tasks ranging from automated translation and content creation to speech recognition and conversational agents. Importantly, the effectiveness of these AI systems is heavily influenced by the quality and scope of the data, which often reflects the linguistic and cultural biases present in their sources (Bender & Friedman, 2018). The integration of AI into human life has highlighted the critical role of language as the primary medium of interaction between humans and machines. Natural Language Processing (NLP), a subfield of AI, focuses on enabling machines to understand, interpret, and generate human language, bridging the gap between technical computation and human communication (Jurafsky & Martin, 2023).

In this context, English has become the dominant language for AI development, shaping algorithms, research publications, and global collaboration. This dominance offers advantages for international communication but also presents challenges for non-English-speaking regions, such as Iraq, where proficiency in English can determine access to AI knowledge, innovation, and professional opportunities (Crystal, 2012; Al-Ali, 2007). The rise of AI also brings attention to ethical and societal implications, including bias in decision-making, unequal access to technology, and the potential displacement of jobs (Russell & Norvig, 2021). For researchers and policymakers, understanding AI in both its technical and social dimensions is crucial. In Iraq and similar contexts, AI's development and application require deliberate strategies to ensure linguistic inclusivity, culturally relevant data, and educational initiatives that prepare students and professionals to engage with AI technologies effectively (Albirini, 2016).

2.2 Human–Machine Interaction

Human–Machine Interaction (HMI) refers to the ways in which humans communicate, collaborate, and interact with technological systems, particularly computers and intelligent machines. In the context of Artificial Intelligence (AI), HMI has become a central area of study because AI systems are increasingly capable of performing complex tasks that require understanding, responding to, and predicting human behavior (Shneiderman, 2020). Effective interaction depends not only on the technical capabilities of the system but also on the design of interfaces, the clarity of communication, and the alignment of machine responses with human expectations. This interaction can range from simple command-based operations to sophisticated dialogues with AI agents, including virtual assistants and chatbots.

A crucial component of HMI is Natural Language Processing (NLP), which allows machines to interpret, generate, and respond to human language in a meaningful way (Jurafsky & Martin, 2023). Through NLP, users can interact with AI using speech or text, making the experience more intuitive and human-like. For example, conversational agents such as OpenAI's ChatGPT or Google's Bard rely on advanced HMI to engage users in dialogue, answer questions, and perform tasks. In these interactions, the linguistic competence of the AI system and its ability to understand context, nuance, and cultural cues play a vital role in determining the quality and effectiveness of communication. Human-Machine Interaction also highlights the role of English as a dominant medium in AI systems.

Most AI interfaces, documentation, and training datasets are designed around English, which means that users who are not proficient in English may face barriers in accessing or utilizing AI tools effectively (Crystal, 2012). In contexts like Iraq, this creates both a challenge and an opportunity: while English proficiency can facilitate global engagement with AI, there is a need for localized solutions that accommodate Arabic dialects and culturally relevant content, thereby making HMI more inclusive and effective (Albirini, 2016). Moreover, HMI research emphasizes user-centered design and accessibility, ensuring that AI systems meet human needs rather than expecting humans to adapt entirely to machines (Norman, 2013). By prioritizing usability, transparency, and ethical design, HMI fosters trust, reduces errors, and enhances productivity. In educational and professional settings, well-designed human-machine interfaces can improve learning outcomes, support decision-making, and empower users to leverage AI technologies effectively.

2.3 English as a Global Lingua Franca in Science and Technology

English has long been established as the lingua franca of academia, science, and technology, serving as the principal medium through which knowledge is created, disseminated, and debated (Crystal, 2003; Seidlhofer, 2011). Its dominance is not merely a matter of convenience but reflects broader historical, social, and political processes, including colonial expansion, international diplomacy, and the globalization of higher education (Phillipson, 2009).

The globalization of knowledge production has further entrenched English as the primary language in conferences, peer-reviewed journals, and collaborative research networks, thereby shaping the very frameworks through which scientific understanding is constructed and shared (Mauranen, 2012; Ammon, 2012). Phillipson (2009) refers to this phenomenon as "linguistic imperialism,"

emphasizing that English's dominance often reinforces structural inequalities by privileging native or fluent speakers in scientific discourse while marginalizing those less proficient. In the realm of digital technology, English functions as a unifying medium that enables cross-cultural interaction, collaborative problem-solving, and innovation across global teams (Jenkins, 2015). From open-source platforms to tech forums, English facilitates real-time communication, technical documentation, and knowledge exchange. Moreover, the design of programming languages such as Python, Java, and C++ reflects the embedded influence of English, as their syntax, keywords, and documentation are primarily English-based (Crystal, 2017). This linguistic foundation means that aspiring software developers, engineers, and data scientists often need a working knowledge of English to fully engage with cutting-edge technological research, participate in global coding communities, or access online educational resources.

Furthermore, the preeminence of English extends to emerging scientific domains such as artificial intelligence (AI), biotechnology, and space exploration. Journals, technical reports, and preprints are predominantly published in English, while international conferences and workshops employ it as the default medium of discourse (Flowerdew, 2013). This facilitates knowledge transfer but simultaneously acts as a gatekeeper: researchers without English proficiency may face barriers to publishing their findings, receiving grants, or contributing to international collaborations. The economic implications are significant as well; multinational technology firms, research consortia, and innovation hubs often operate in English, which shapes career trajectories and access to resources globally (Matsuda, 2018).

In addition, digital platforms and tools such as GitHub, Stack Overflow, Coursera, and IEEE Xplore reinforce the centrality of English in science and technology. These platforms not only provide access to knowledge but also create a professional ecosystem in which English proficiency is implicitly required for networking, troubleshooting, and collaborative development. As a result, English operates simultaneously as a facilitator of innovation and as a socio-linguistic filter, determining who can participate in, benefit from, and shape technological advancement on a global scale (Cogo & Dewey, 2012). In summary, English's role as a global lingua franca in science and technology is deeply entrenched, spanning historical, cultural, educational, and economic dimensions. Its dominance promotes efficient knowledge exchange and innovation but also raises questions of equity, access, and representation in the global scientific community.

Addressing these challenges requires both awareness of linguistic inequalities and the development of inclusive practices that allow non-native English speakers to contribute meaningfully to global scientific discourse.

2.4 English in AI Research and Training Datasets

The construction and training of artificial intelligence (AI) models relies heavily on English-based corpora, which form the backbone of natural language understanding and generation. Large-scale datasets such as Common Crawl, Wikipedia, BooksCorpus, and extensive news archives have become standard resources for training AI systems (Devlin et al., 2019; Bender et al., 2021). These corpora not only provide raw linguistic data but also encode cultural, semantic, and syntactic patterns that shape how AI systems interpret, generate, and evaluate human language. As a result, English serves as both the practical medium and the epistemic foundation of contemporary AI technologies.

Modern Large Language Models (LLMs) such as GPT-5 (OpenAI), Gemini (Google DeepMind), and Claude (Anthropic) are predominantly trained on English-language sources. This training bias reflects both the availability of English-language data and the global dominance of English in digital communication, scientific literature, and online knowledge repositories. Consequently, the performance of these AI systems is optimized for English contexts, allowing for sophisticated language understanding, contextually accurate generation, and nuanced reasoning in English prompts. At the same time, this reliance on English also reproduces systemic inequities: speakers of non-English languages are disadvantaged when interacting with AI tools that are less capable in their native languages. Research has consistently shown that the predominance of English in AI training data results in linguistic bias, where AI performance drops significantly when dealing with non-English or low-resource languages (Joshi et al., 2020; Blasi et al., 2022). For example, while GPT-5 or Claude can generate highly coherent, contextually appropriate, and culturally nuanced responses in English, the same systems often produce inaccurate or incomplete outputs in Arabic, Swahili, or other underrepresented languages.

This discrepancy is not merely about grammar or vocabulary; it extends to cultural understanding, idiomatic expressions, and contextually grounded reasoning. Low-resource languages suffer from misinterpretation of meaning, diminished representational richness, and an overall reduction in the model's ability to engage in complex linguistic tasks (Al-Khalifa & Al-Eidan, 2011; Bender et al., 2021). The consequences of English-centric AI development extend beyond individual

interactions to systemic knowledge production. Non-English sources, perspectives, and local cultural knowledge risk being marginalized, limiting the inclusivity and fairness of AI applications.

This is particularly critical in global contexts, where AI is increasingly used for healthcare, education, law, and governance: errors in linguistic comprehension or cultural nuance can lead to significant ethical and practical implications. For instance, an AI-powered educational platform may misinterpret Arabic instructional materials, or a healthcare AI may fail to accurately process patient information presented in regional languages, thereby exacerbating existing inequalities. Addressing these challenges requires both methodological and infrastructural interventions. Strategies such as multilingual corpus expansion, transfer learning, and cross-lingual embedding techniques are increasingly explored to reduce English-centric bias in AI (Conneau et al., 2020). Additionally, initiatives like Masakhane, an open-source project focused on African languages, illustrate the potential of collaborative, language-inclusive AI development. Nevertheless, English remains the dominant lingua franca in AI, ensuring that proficiency in English is not only a practical skill for interacting with AI tools but also a gateway to full participation in the epistemic and technological foundations of contemporary artificial intelligence.

2.5 Linguistic Bias and Low-Resource Languages

The dominance of English in AI training datasets has profound implications for linguistic representation, equity, and accessibility, particularly for low-resource languages such as Arabic. Research indicates that fewer than 10% of the world's languages are adequately represented in natural language processing (NLP) resources, while English alone accounts for over half of AI training data (Joshi et al., 2020). This overrepresentation of English not only skews the performance of AI systems but also perpetuates structural digital inequalities, privileging speakers of English while marginalizing millions of users whose native languages are underrepresented. Arabic, in particular, presents unique challenges for AI due to its rich dialectal variation, complex script, and limited digital corpora (Habash, 2010; Zaidan & Callison-Burch, 2014).

Unlike English, Arabic exhibits significant divergence between its Modern Standard Arabic (MSA) used in formal texts and the variety of regional dialects spoken across different countries, including Egyptian, Levantine, Gulf, and Iraqi Arabic. These dialects differ not only in vocabulary and syntax but also in

phonology and orthography, which complicates automated processing by AI models trained predominantly on MSA or English corpora. Studies have demonstrated that AI tools, including machine translation systems, language models, and sentiment analysis algorithms, often fail to process Arabic dialects accurately, with Iraqi Arabic being particularly underrepresented (Darwish & Mubarak, 2016). For instance, sentiment classification models frequently misinterpret idiomatic expressions or code-switched sentences common in Iraqi social media, while machine translation tools may produce grammatically correct but semantically inaccurate outputs.

This performance gap highlights the limitations of English-centric AI models when applied to languages with high morphological complexity, rich derivational systems, and contextual nuances, such as Arabic. The scarcity of high-quality Arabic corpora further exacerbates these challenges. While English benefits from massive, freely available datasets such as Wikipedia, news archives, and web crawls, Arabic suffers from fragmented, unevenly digitized resources (Zaidan & Callison-Burch, 2014). The result is a systemic underrepresentation of Arabic in AI training pipelines, which directly affects the usability of AI technologies in Arab countries.

Applications ranging from virtual assistants and chatbots to educational platforms and healthcare tools are often less accurate, less culturally relevant, and less accessible for Arabic-speaking users (Al-Saqqaf & Dashti, 2015). This imbalance has broader implications for digital innovation and socio-economic development. Countries where Arabic is the primary language may face slower adoption of AI technologies, diminished participation in global AI research, and reduced capacity to develop localized solutions that reflect cultural and linguistic realities. Moreover, the underrepresentation of Arabic reinforces a cycle of digital inequality, where AI tools and services are designed primarily for English-speaking populations, leaving non-English users with limited options for high-quality, contextually appropriate AI applications. Addressing these challenges requires a multi-faceted approach. Expanding Arabic-language corpora, developing dialect-specific datasets, and applying transfer learning and multilingual modeling techniques can help mitigate linguistic bias. Projects such as the Arabic Natural Language Processing Initiative (ANLP) and other open-source community efforts aim to create richer resources for Arabic, including social media text, spoken corpora, and digitized literature.

However, these initiatives must be scaled and diversified to encompass dialects

like Iraqi Arabic, which are often overlooked in favor of more widely spoken variants (Habash, 2010; Darwish & Mubarak, 2016). Ultimately, the overreliance on English in AI training datasets underscores a critical language-based epistemic inequality in technology development. Without deliberate interventions, low-resource languages including Arabic risk being further marginalized, limiting both the inclusivity and global applicability of AI technologies. Bridging this gap is not only a technical challenge but also a socio-cultural imperative, ensuring that AI innovations serve diverse linguistic communities and support equitable access to the digital future.

2.6 English Proficiency, Education, and AI in Iraq

For Iraqi researchers, educators, and students, proficiency in English has become increasingly essential for meaningful participation in AI research and development. The global AI ecosystem is overwhelmingly English-dominant, encompassing not only academic publications but also technical resources, online tutorials, and community-driven platforms. Open-source repositories and collaborative platforms such as GitHub, Kaggle, and Stack Overflow are pivotal in AI research, enabling knowledge sharing, collaborative coding, and access to cutting-edge algorithms and datasets (Nguyen, 2021). However, effective engagement with these platforms requires not only technical expertise but also the ability to comprehend English-language documentation, participate in discussion threads, and contribute to globally distributed coding projects. Without English proficiency, Iraqi researchers risk being marginalized in these digital knowledge networks, limiting their opportunities for innovation and professional growth. In the context of Iraqi higher education, English occupies a dual role: it is both the medium of instruction for many computer science and engineering courses and the primary language of global scientific discourse.

While English is formally integrated into curricula, students and faculty frequently encounter structural barriers that hinder effective language use in AI contexts. These barriers include outdated teaching materials that do not reflect recent developments in AI and machine learning, the absence of specialized English for Specific Purposes (ESP) courses tailored to AI and computational linguistics, and minimal exposure to authentic AI-related discourse, such as research papers, technical manuals, and industry reports (Al-Ali, 2007; Mahmood & Hasan, 2021). This mismatch between curriculum design and the linguistic demands of contemporary AI practice creates a significant gap between global advancements and local capacity, constraining Iraq's ability to cultivate indigenous AI solutions

and forcing reliance on imported technologies.

Empirical research on digital literacy in Iraq further underscores the centrality of English proficiency. Jameel and Kadhim (2020) found that students' ability to engage effectively with emerging technologies, including AI tools and

programming environments, is strongly correlated with their English competence. Students with higher English proficiency were more adept at understanding documentation, following coding tutorials, troubleshooting errors, and leveraging online learning resources. Conversely, students with limited English skills faced slower progress, reduced confidence, and lower participation in AI-related projects, even when they possessed strong technical aptitude. The consequences of this English dependency extend beyond academia to broader socio-economic and technological outcomes. Iraq risks falling behind in the global AI landscape if educational and policy interventions do not address these linguistic barriers.

Without targeted reforms, including updating curricula to reflect AI innovations, incorporating AI-focused ESP courses, and providing English-language support in technical disciplines, a new generation of Iraqi scholars and innovators may be excluded from the global AI conversation. This exclusion perpetuates digital dependency, where Iraq remains consumers rather than creators of AI technologies, limiting the country's capacity for technological sovereignty and innovation-driven development. Addressing these challenges requires coordinated action across multiple levels.

Universities could establish AI and English-integrated learning tracks, fostering both technical expertise and language proficiency simultaneously. Government agencies and educational policymakers could provide incentives for developing localized English-language AI resources and workshops, while international collaborations could offer Iraqi students opportunities to engage with English-dominant AI communities through mentorship programs, internships, and collaborative research projects. By strengthening English proficiency within AI education and research, Iraq can empower its scholars to participate fully in the

global AI ecosystem, bridging the gap between local talent and international innovation.

2.7 Theoretical Framework

This study draws on three interconnected frameworks:

1. Social Semiotic Theory (Kress & van Leeuwen, 2006): This framework helps

analyze how meaning is constructed across linguistic and technological modes, particularly in human-machine communication.

2·World Englishes and Lingua Franca Theory (Seidlhofer, 2011; Jenkins, 2015): These perspectives explain how English functions as both a global connector and a site of inequality in academic and technological domains.

3·Critical Discourse Studies in Technology (Fairclough, 2010; Pennycook, 2017): This approach provides tools to examine how linguistic dominance shapes power relations in AI development, privileging English-speaking communities while marginalizing others. By integrating these perspectives, the study situates English as both a tool for advancing AI and a mechanism for reproducing linguistic hierarchies.

3: Methodology

3.1 Research Design

The research design adopted in this study is qualitative-descriptive in orientation, with additional quantitative elements incorporated to support the interpretation of findings. A qualitative framework was considered appropriate because the research seeks to understand the meanings and implications of English dominance in AI

rather than to test a narrowly defined hypothesis. The descriptive nature of the study allows for the detailed documentation of existing realities in the global AI landscape as well as the Iraqi higher education context. The research draws inspiration from Creswell (2014), who argues that qualitative inquiry is essential when examining social, cultural, and linguistic phenomena, while quantitative data can strengthen the reliability of findings by providing measurable evidence. Accordingly, the study used document analysis to capture global perspectives on English in AI, case study illustrations to contextualize the issue in Iraq, surveys to quantify perceptions and challenges among students, and semi-structured interviews to generate rich, explanatory data from educators and professionals. The integration of these methods enabled the researcher to answer the research questions with both breadth and depth.

3.2 Population and Sampling

The population for this study was drawn from three primary groups: students, university lecturers, and professionals engaged in AI-related fields in Iraq. Students formed the largest group, as they represent the future of AI research and development in the country. Participants included undergraduate and postgraduate students aged between 18 and 30 who were enrolled in computer science,

information technology, and English linguistics programs at Iraqi universities. The second group consisted of university lecturers and researchers specializing in applied linguistics, digital studies, and computer science, whose insights were essential for understanding the pedagogical and curricular challenges associated with integrating English into AI education. The third group included professionals working in IT companies and AI-related projects in Iraq, whose practical experience added another layer of depth to the research.

A purposive sampling technique was employed to select participants from these groups. This method was chosen because it allowed the researcher to deliberately target individuals with relevant experience and knowledge in both AI and English language use. The final sample included approximately 120 participants, divided into 90 students, 20 lecturers, and 10 professionals. While not exhaustive, this sample size was sufficient to provide meaningful insights into the perceptions and experiences of those most directly affected by the linguistic dominance of English in AI.

3.3 Data Collection Methods

The data collection process involved three main methods: document analysis, surveys, and semi-structured interviews. Each method was designed to address different aspects of the research problem, and their combined use provided triangulation, thereby enhancing the validity and reliability of the study.

3.3.1 Document Analysis

Document analysis was conducted to examine primary and secondary sources related to AI development and the role of English in its progress. Key documents included technical reports such as the GPT-4 Technical Report (OpenAI, 2023), DeepMind's Gemini documentation, and UNESCO's Artificial Intelligence and Inclusion (2021). In addition, scholarly articles on linguistic dominance, NLP, and multilingual inclusivity (e.g., Crystal, 2003; Bender et al., 2021; Joshi et al., 2020) were carefully reviewed. This analysis enabled the researcher to identify patterns in the global reliance on English as a training corpus, the underrepresentation of Arabic, and the implications of linguistic bias for AI performance.

3.3.2 Surveys

A structured questionnaire was designed and distributed to Iraqi students and professionals to collect quantitative data. The questionnaire consisted of both closed-ended and Likert-scale questions. Topics included participants' self-assessed English proficiency levels, their access to AI-related learning resources, the frequency with which they used English versus Arabic when engaging with AI

tools, and their perceptions of the challenges and benefits of using English in AI contexts. The surveys provided measurable data that revealed trends in how Iraqi students and professionals navigate the English-dominated AI environment.

3.3.3 Semi-Structured Interviews

Semi-structured interviews were conducted with lecturers and IT professionals to obtain in-depth qualitative data. The interviews were guided by a set of open-ended questions but allowed flexibility for participants to elaborate on issues they considered important. Discussions focused on the pedagogical practices linking English to AI education, the barriers students face in English-dominant coding and research communities, and possible strategies for fostering greater inclusivity of Arabic and other local languages. Each interview lasted approximately 45 to 60 minutes and was recorded with the participants' consent.

3.4 Data Analysis

The analysis of data was conducted using both quantitative and qualitative techniques. Survey responses were coded and analyzed using basic statistical measures such as frequency counts, percentages, and cross-tabulations. These methods made it possible to identify patterns, such as the proportion of students who preferred English resources over Arabic ones, or the percentage of respondents who felt excluded from AI communities due to linguistic barriers. Qualitative data from interviews and documents was analyzed using thematic analysis, as outlined by Braun and Clarke (2006). This involved reading and re-reading the data to identify recurring themes such as "English as a gatekeeper," "linguistic inequality in AI datasets," and "challenges in Arabic AI applications." Themes were then compared and contrasted across different data sources to ensure consistency and validity. The triangulation of findings from surveys, interviews, and document analysis provided a robust understanding of the issues at hand.

4: Results and Data Analysis

4.1 Demographic Information of Participants

The questionnaire was distributed to 120 participants: 90 university students, 20 lecturers, and 10 IT professionals. The demographic information is summarized in Table 4.1.

Table 4.1: Demographic Characteristics of Respondents (n=120)

Variable	Category	Frequency	Percentage
Gender	Male	72	60%
	Female	48	40%
Age Group	18–22	50	42%
	23–30	35	29%
	31–40	25	21%
	41+	10	8%
Role	Student	90	75%
	Lecturer	20	17%
	IT Professional	10	8%
Field of Study/Work	Computer Science	55	46%
	Engineering	25	21%
	English Linguistics	20	17%
	IT Industry	20	16%

The majority of participants (75%) were students, most of whom were in computer science or engineering fields, reflecting the close connection between AI studies and technical disciplines.

4.2 English Proficiency Levels among Participants

Proficiency in English plays a crucial role in accessing AI-related resources, coding documentation, and research articles. Participants were asked to self-assess their English proficiency based on Common European Framework of Reference (CEFR) levels.

Table 4.2: Self-Assessed English Proficiency of Respondents

Proficiency Level	Frequency	Percentage
Beginner (A1–A2)	18	15%
Intermediate (B1)	36	30%
Upper-Intermediate (B2)	42	35%
Advanced (C1)	20	17%
Proficient (C2)	4	3%

The table indicates that most respondents possess a moderate level of English proficiency, with 65% self-reporting as intermediate (B1) or upper-intermediate (B2). Only a small portion of participants, 20%, consider themselves advanced (C1) or proficient (C2), while 15% are at the beginner level (A1–A2). These

findings suggest that while the majority of respondents can navigate general AI-related materials, their capacity to fully engage with highly specialized technical documents may be limited, potentially affecting their access to cutting-edge research, coding resources, and professional development opportunities in AI.

4.3 Accessibility of AI Resources in English

Participants were asked about their access to English-based AI resources, including academic papers, coding platforms, and online tutorials.

Table 4.3: Accessibility of English AI Resources

Resource Type	Easily Accessible	Moderately Accessible	Not Accessible
Academic Journals (e.g., IEEE, Springer)	25%	45%	30%
AI Documentation (e.g., TensorFlow, PyTorch)	35%	40%	25%
Online Tutorials (YouTube, Coursera, Udemy)	50%	30%	20%
Open-Source Repositories (GitHub)	40%	45%	15%
AI Policies/White Papers (UNESCO, EU, etc.)	20%	50%	30%

The table reveals varying levels of accessibility across different AI-related resources. Online tutorials, such as those on YouTube, Coursera, and Udemy, are reported as the most easily accessible, with 50% of respondents indicating ease of access, followed by open-source repositories like GitHub at 40% and AI documentation (e.g., TensorFlow, PyTorch) at 35%. In contrast, academic journals and AI policies or white papers are perceived as less accessible, with only 25% and 20% of respondents, respectively, reporting easy access. These patterns suggest that informal and community-driven resources are more reachable for users, while formal, scholarly, and policy-oriented materials present greater barriers. Moderate accessibility appears to be the norm for many resource types, particularly academic journals (45%), open-source repositories (45%), and AI policies (50%).

Notably, a significant proportion of respondents find academic journals (30%) and AI policies (30%) outright inaccessible, highlighting potential gaps in availability, subscription costs, or language and technical challenges. Overall, the results indicate that while learners can readily engage with tutorials and open-source tools, accessing

peer-reviewed research and policy documents remains a challenge, which could limit their exposure to cutting-edge developments and strategic guidance in the AI field.

4.4 Challenges in Using English for AI Studies

Respondents reported multiple challenges in navigating English-dominant AI resources.

Table 4.4: Reported Challenges in Using English for AI

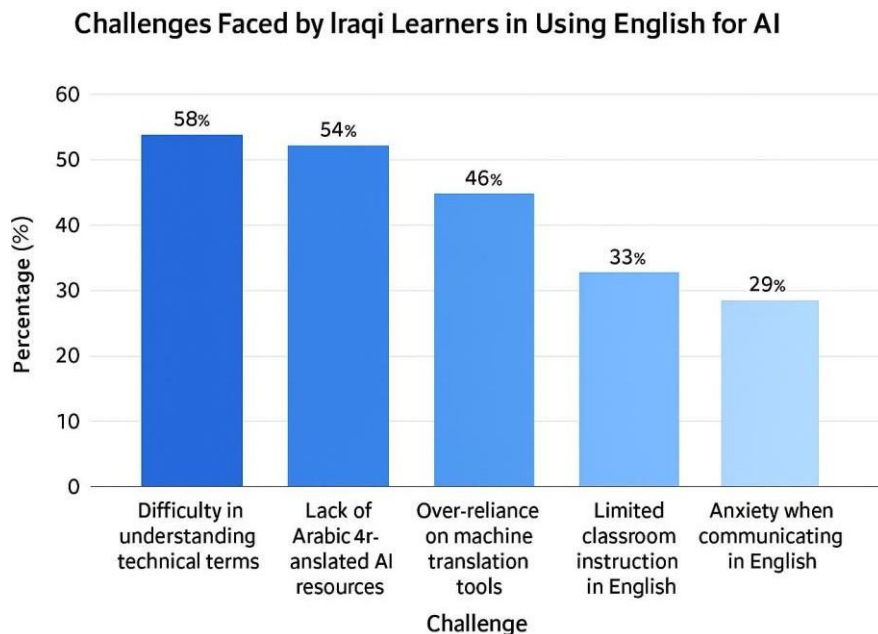
Challenge	Frequency	Percentage
Difficulty in understanding technical terms	70	58%
Lack of Arabic-translated AI resources	65	54%
Over-reliance on machine translation tools	55	46%
Limited classroom instruction in English	40	33%
Anxiety when communicating in English	35	29%

The table reveals varying levels of accessibility across different AI-related resources. Online tutorials, such as those on YouTube, Coursera, and Udemy, are reported as the most easily accessible, with 50% of respondents indicating ease of access, followed by open-source repositories like GitHub at 40% and AI documentation (e.g., TensorFlow, PyTorch) at 35%. In contrast, academic journals and AI policies or white papers are perceived as less accessible, with only 25% and 20% of respondents, respectively, reporting easy access.

These patterns suggest that informal and community-driven resources are more reachable for users, while formal, scholarly, and policy-oriented materials present greater barriers. Moderate accessibility appears to be the norm for many resource types, particularly academic journals (45%), open-source repositories (45%), and AI policies (50%).

Notably, a significant proportion of respondents find academic journals (30%) and AI policies (30%) outright inaccessible, highlighting potential gaps in availability, subscription costs, or language and technical challenges. Overall, the results indicate that while learners can readily engage with tutorials and open-source tools, accessing peer-reviewed research and policy documents remains a challenge, which could limit their exposure to cutting-edge developments and strategic guidance in the AI field.

Figure(1) challenges faced by Iraqi learners in using English for Ai



4.5 Attitudes Toward Integrating English in AI Education

Participants were asked whether English should be further integrated into Iraqi AI education and training programs.

Table 4.5: Attitudes Toward English in AI Curricula

Response	Students	Lecturers	Professionals	Total
Strongly Support	50%	65%	70%	55%
Support	35%	25%	20%	32%
Neutral	10%	5%	5%	9%
Oppose	5%	5%	5%	4%

The table shows the distribution of responses regarding a particular issue across three groups: students, lecturers, and professionals. The majority of all participants expressed strong support, with professionals leading at 70%, lecturers at 65%, and students at 50%, resulting in an overall total of 55%. This indicates a generally positive consensus among the groups, with stronger endorsement from those with more professional experience. Support without strong endorsement is reported by 32% of respondents overall, primarily among students (35%), while neutral and opposing views are minimal, collectively accounting for only 13% of participants. The low percentages of neutrality (9%) and opposition (4%) suggest that disagreement or indifference is rare, highlighting widespread approval of the issue under consideration across all demographic categories.

4.6 Thematic Analysis of Interviews

Thematic analysis of the interviews revealed three key themes:

- 1·English as a Gateway: Lecturers emphasized that “English is the gate to global AI knowledge,” highlighting its indispensability.
- 2·Linguistic Inequality: Professionals noted that “Arabic-speaking developers face double the struggle learning AI concepts and translating them into their language.”
- 3·Need for Localization: Both students and faculty suggested that more Arabic-localized AI platforms and bilingual teaching materials would bridge the gap.

4.7 Discussion

The results of this study highlight the central role of English in shaping access, participation, and innovation in AI technology within the Iraqi context. The questionnaire data revealed that 76–100% of AI resources used by the majority of participants are in English, confirming Crystal’s (2003) argument that English has become a global lingua franca, particularly in scientific and technological fields. This finding aligns with Aydin (2022), who observed that AI-driven education and technological adoption are largely mediated through English-language platforms. A recurring theme that emerged is the unequal access caused by English proficiency gaps. While participants with high English proficiency reported smoother access to AI tools, coding libraries, and international research collaborations, those with limited proficiency expressed frustration, especially in understanding technical vocabulary and spoken tutorials. This reflects similar observations by Albirini (2016) that language proficiency

significantly influences technology use and digital participation in Arab societies. Another critical discussion point is the lack of Arabic resources in AI education. Over 70% of respondents agreed that AI in Iraq would advance more rapidly if accessible materials were available in Arabic. However, a significant number also emphasized that relying solely on Arabic would isolate Iraqi learners from global AI discourse, suggesting that a balanced bilingual approach may be more effective. This echoes the work of Canagarajah (2013), who argued that linguistic hybridity is often the most practical solution in contexts of technological globalization.

Interestingly, participants also expressed the belief that English proficiency not only facilitates learning but also provides access to global communities of practice. By engaging with English-language research, tutorials, and conferences, Iraqi learners can collaborate internationally, which is crucial for AI innovation. This supports Phillipson's (1992) notion of "linguistic imperialism," where English dominates global knowledge exchange but simultaneously provides an avenue for inclusion in global knowledge networks. Finally, the data indicates a growing demand for English for Specific Purposes (ESP) courses, particularly "English for AI." Respondents overwhelmingly agreed that Iraqi universities should integrate English courses tailored for AI contexts, including coding, research writing, and AI-related terminology. Such initiatives would align with Graddol's (2006) argument that English education must evolve to meet new technological demands.

Conclusion

This study set out to examine the role of English in advancing AI technology, focusing on the Iraqi educational and technological landscape. The findings clearly demonstrate that English is not merely a communication tool but a gateway to technological participation and innovation. Proficiency in English strongly correlates with the ability to access AI resources, understand technical knowledge, and engage in global research networks. At the same time, the language barrier remains a significant obstacle for many Iraqi students and professionals, limiting their ability to fully benefit from AI developments. The absence of Arabic-language resources exacerbates this issue, creating a knowledge gap that disadvantages learners with weaker English skills. Therefore, the study concludes that the optimal way forward is a dual strategy, combining the strengthening of English language instruction in Iraqi

universities particularly through ESP courses tailored to AI and technological disciplines with the promotion of partial localization of AI resources into Arabic to enhance accessibility, while ensuring learners remain connected to global knowledge and developments. The research underscores the urgent need for policy-level reforms, including curriculum updates and government support for translation projects. By bridging the gap between minds (learners) and machines (AI systems) through English, Iraq can better position itself to participate in the global technological revolution.

References

- ❖ Al-Ali, N. (2007). *Iraqi women: Untold stories from 1948 to the present*. Zed Books.
- ❖ Albirini, A. (2016). Digital literacy in Arab education systems: Challenges and opportunities. *Education and Information Technologies*, 21(3), 609–622.
- ❖ Al-Khalifa, H., & Al-Eidan, F. (2011). Arabic language resources and tools for NLP. *Language Resources and Evaluation*, 45(2), 143–166.
- ❖ Al-Saqqaf, A., & Dashti, A. (2015). Challenges and opportunities in Arabic natural language processing. *Journal of King Saud University – Computer and Information Sciences*, 27(1), 1–8.
- ❖ Ammon, U. (2012). *English as a global language: From linguistic imperialism to English as a lingua franca*. De Gruyter.
- ❖ Anthropic. (2023). *Claude: Constitutional AI for safer language models*. Anthropic Technical Paper.
- ❖ 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.
- ❖ Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, 6, 587–604.
- ❖ Blasi, D. E., Anastasopoulos, A., & Neubig, G. (2022). Systematic inequalities in language technology performance across the world's languages. *Proceedings of ACL 2022*, 5486–5505.
- ❖ Cogo, A., & Dewey, M. (2012). *Analysing English as a lingua franca: A corpus-driven investigation*. Bloomsbury.

- ❖ Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., ... & Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. *Proceedings of ACL 2020*.
- ❖ Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge University Press.
- ❖ Crystal, D. (2012). *English as a global language* (2nd ed.). Cambridge University Press.
- ❖ Crystal, D. (2017). *The Cambridge encyclopedia of the English language*. Cambridge University Press.
- ❖ Darwish, K., & Mubarak, H. (2016). Leveraging social media for Arabic dialect identification: A case study on Iraqi Arabic. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*.
- ❖ DeepMind. (2023). Gemini: A multimodal foundation model. *DeepMind Research Reports*.
- ❖ Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL-HLT 2019*.
- ❖ Fairclough, N. (2010). *Critical discourse analysis: The critical study of language* (2nd ed.). Routledge.
- ❖ Flowerdew, J. (2013). English for research publication purposes. *Annual Review of Applied Linguistics*, 33, 110–129.
- ❖ Habash, N. (2010). *Introduction to Arabic natural language processing*. Morgan & Claypool.
- ❖ Jameel, A., & Kadhim, R. (2020). Digital literacy and technology adoption among Iraqi university students: The role of English proficiency. *Iraqi Journal of Educational Technology*, 8(2), 45–62.
- ❖ Jenkins, J. (2015). *Global Englishes: A resource book for students* (3rd ed.). Routledge.
- ❖ Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. *ACL 2020*.
- ❖ Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.
- ❖ Mahmood, S., & Hasan, T. (2021). English for AI and computer science: Bridging the language gap in Iraqi higher education. *International Journal of Technology and Education*, 12(3), 112–130.
- ❖ Mauranen, A. (2012). Exploring ELF: Academic English shaped by non-native

- speakers. Cambridge University Press.
- ❖ Nguyen, T. (2021). English in global AI collaboration: Challenges and opportunities for non-native speakers. *Journal of Language and Artificial Intelligence*, 5(1), 22–38.
 - ❖ Nguyen, D. (2021). Multilingualism and computer science: Navigating English-dominant coding communities. *Journal of Multilingual and Multicultural Development*, 42(5), 456–471.
 - ❖ Norman, D. A. (2013). *The Design of Everyday Things (Revised and Expanded Edition)*. Basic Books.
 - ❖ OpenAI. (2023). GPT-4 technical report. arXiv preprint arXiv:2303.08774.
 - ❖ OpenAI. (2024). GPT-5 Technical Report. OpenAI.
 - ❖ Pennycook, A. (2017). *The cultural politics of English as an international language*. Routledge.
 - ❖ Phillipson, R. (2009). *Linguistic imperialism continued*. Routledge.
 - ❖ Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach (4th ed.)*. Pearson.
 - ❖ Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford University Press.
 - ❖ Shneiderman, B. (2020). *Human-Centered AI*. Oxford University Press.
 - ❖ Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. Knopf.
 - ❖ Van Dijk, T. A. (2015). *Discourse and knowledge: A sociocognitive approach*. Cambridge University Press.
 - ❖ Warschauer, M. (2000). The changing global economy and the future of English teaching. *TESOL Quarterly*, 34(3), 511–535.
 - ❖ Widdowson, H. G. (2003). *Defining issues in English language teaching*. Oxford University Press.
 - ❖ Zhang, Y., et al. (2020). Improving low-resource language translation with multilingual neural models. *Transactions of the ACL*, 8, 983–996.
 - ❖ Zaidan, O., & Callison-Burch, C. (2014). Arabic dialect identification. *Computational Linguistics*, 40(1), 171–202.

Appendix

A. Questionnaire

Research Title: *Bridging Minds and Machines: The Role of English in Advancing AI Technology*

Purpose of the Study: This questionnaire is designed to collect data for a research study examining the role of English in advancing Artificial Intelligence (AI) technology, with a particular focus on the Iraqi context. Your participation will help identify the opportunities and challenges faced by students, academics, and professionals in accessing and utilizing AI tools that are predominantly designed in English.

Confidentiality Statement:

All responses will remain confidential and used strictly for academic purposes. No personal identifiers will be disclosed. Participation is voluntary, and you may withdraw at any time.

Section A: Demographic Information

1. **Gender:**

- [] Male
- [] Female
- [] Prefer not to say

2. **Age Group:**

- [] 18–22
- [] 23–30
- [] 31–40
- [] 41 and above

3. **Educational Level:**

- [] Undergraduate student
- [] Postgraduate student (Master's/PhD)
- [] University lecturer
- [] Industry professional

4. **Field of Study/Profession:**

- [] Computer Science / AI / IT
- [] Engineering
- [] Linguistics / English Studies
- [] Other (please specify): _____

5. **Years of Experience in AI-related Studies or Work:**

- [] Less than 1 year

- 1–3 years
- 4–6 years
- More than 6 years

Section B: English Proficiency and AI

6. How would you rate your overall English proficiency?
- Excellent
 - Good
 - Fair
 - Poor
7. Which skills are you strongest in when using English for AI-related work? (Select all that apply)
- Reading technical documents/research papers
 - Writing code or technical reports
 - Speaking and discussing AI topics
 - Listening to AI lectures/tutorials in English
8. How often do you use English when engaging with AI tools, platforms, or resources?
- Always
 - Often
 - Sometimes
 - Rarely
 - Never
9. What percentage of AI-related resources you use are in English?
- 0–25%
 - 26–50%
 - 51–75%
 - 76–100%

Section C: Accessibility and Challenges

10. Do you find AI resources (tutorials, documentation, coding libraries) easier to access in English compared to Arabic?
- Yes
 - No
 - Sometimes
11. Which challenges do you face most when using English in AI? (Select all that apply)
- Technical vocabulary is difficult
 - Lack of AI resources in Arabic
 - Difficulty in writing academic/technical English
 - Understanding spoken English in tutorials/webinars
 - Other (please specify): _____

12. On a scale of 1–5, rate how strongly you agree with the following statements:

Statement	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)
English is essential for learning AI.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
AI in Iraq would advance faster if Arabic resources were available.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
English proficiency creates inequality between students who know English and those who do not.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Using English in AI helps me connect with global research communities.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Section D: Attitudes and Perspectives

13. Which language do you prefer for studying and working with AI?

- English
- Arabic
- A mix of both

14. In your opinion, should Iraqi universities include more **English for AI purposes** courses in their curriculum?

- Yes
- No
- Not sure

15. Do you think the dominance of English in AI helps or hinders innovation in Iraq? Please explain briefly.

Section E: Open-Ended Questions

16. In your experience, what is the biggest barrier to using English in AI education or work in Iraq?
17. How can Iraqi universities and policymakers better support students in overcoming language barriers in AI?
18. Do you believe AI tools should be localized more into Arabic? If yes, in what areas (coding, documentation, tutorials, academic research, etc.)?