

Artificial Intelligence and the Evolution of English Vocabulary: A Corpus-Based Study

Asst. Inst. Rebin Bahroz Amin

Department of English, College of Education for Human Sciences,
University of Diyala, Diyala, Iraq
rebinzangana44@gmail.com

الذكاء الاصطناعي وتطور المفردات الانجليزية: دراسة قائمة على مجموعة من النصوص

م.م. ريبين بهروز امين

قسم اللغة الانكليزية / كلية التربية للعلوم الإنسانية / جامعة ديالى / ديالى / العراق

Abstract

The study examines the development and use of Artificial Intelligence (AI) terminology within the leading English language Corpora from 1990 to 2023. By analyzing data from the corpus of contemporary American English (Coca), Google Books Ngram Viewer, and now corpus, deep discovery of, first appearance, intelligence, and Chatbot. The findings reveal a significant increase in the use of these conditions over time, reflecting rapid technological progress and increasing social interest in AI. Additionally, emerging words such as Prompt Engineering and AI hallucination highlight linguistic and ideological development in the field. This study underlines the interdisciplinary nature of the AI discourse by incorporating technical, moral and social dimensions, and emphasizes the importance of monitoring language change to better understand the impact of AI on communication and culture. **Keywords:** Artificial Intelligence, AI Terminology, Besides some Emerging Terms, Deep Learning, Chatbot.

المخلص

تدرس هذه الدراسة تطور واستخدام مصطلحات الذكاء الاصطناعي ضمن مجموعات نصوص اللغة الإنجليزية الرائدة من عام ١٩٩٠ إلى عام ٢٠٢٣. ومن خلال تحليل البيانات المستمدة من مجموعة نصوص اللغة الإنجليزية الأمريكية المعاصرة، ومجموعة النصوص الحالية والذكاء والظهور الاول و الاكتشاف العميق وروبوتات الدردشة. تكشف الاستنتاجات عن زيادة ملحوظة في استخدام هذه المصطلحات بمرور الوقت، مما يعكس التقدم التكنولوجي السريع والاهتمام الاجتماعي المتزايد بالذكاء الاصطناعي. بالإضافة إلى ذلك، تسلط الكلمات الناشئة مثل الهندسة السريعة وهلوسة الذكاء الاصطناعي الضوء على التطور اللغوي والأيدولوجي في هذا المجال. تؤكد هذه الدراسة على الطبيعة متعددة التخصصات لخطاب الذكاء الاصطناعي من خلال دمج الأبعاد التقنية والأخلاقية والاجتماعية، وتؤكد على أهمية رصد تغير اللغة لفهم تأثير الذكاء الاصطناعي على التواصل والثقافة بشكل أفضل. **الكلمات المفتاحية:** الذكاء الاصطناعي، مصطلحات الذكاء الاصطناعي، بالإضافة الى بعض المصطلحات الناشئة، التعلم العميق، روبوت الدردشة.

1.Introduction

Artificial Intelligence (AI) has emerged as one of the most transformational techniques of the 21st century, influencing various aspects of society, technology and communication. As AI moves forward, the language is used to describe its concepts, applications and implications. The study of AI-related vocabulary within linguistic Corpora provides valuable insight on how this rapidly developed area is reflected in contemporary discourse, revelations a change in public awareness, scientific attention and cultural approach. The purpose of this research is to check the frequency, emergence and relevant use of major AI words in the English language Corpora on a large scale, such as contemporary American English (Coca), Google Books Ngram Viewer and now Corpus of Corpus. By analyzing these corpora spread over several decades, the study has detected the historical development of AI terminology and highlighted the new vocabulary that has appeared with recent technical

successes. Understanding linguistic trends in AI vocabulary not only contributes to the fields of computational linguistics and corpus studies, but also informs teachers, policy makers and technologists how AI allows concepts to language and society. In addition, searching for embedded moral, social and cultural dimensions in the AI discourse is necessary for the overall understanding of the role of AI beyond purely technical narratives. This introduction determines the stage for AI-related words, their statistical prominence, collage patterns, and a detailed discovery of socio-cultural implications, which contributes to the widespread understanding of the integration of AI in modern English language and idea.

2.Literature ReviewThe study of Artificial Intelligence (AI) terminology has attracted attention as the area of AI develops rapidly. Many researchers have discovered how the AI-related vocabulary emerges, changes, and reflects comprehensive technical and social development. Mikelsonienne and Motiejūnienė (2021) analyzed the large linguistic corpora and found a significant increase in words such as "machine learning" and "deep learning" after the start of the 2000s, which is correlated with progress in computational power and explosion of large data. His findings suggest that linguistic trends reflect technical revolutions, confirming the dynamic relations between language and innovation. Similarly, Uday and Reddy (2024) recently focused on Corpora that to show that the AI vocabulary is becoming increasingly built in educational and public domains not only in the technical discourse but also in educational and public domains. Their job highlights the impact of AI growth on language teaching and course development, emphasizes the need to adapt linguistic resources for emerging AI concepts. The cultural and moral dimensions of the use of AI language have also been studied extensively. Ormanova et al. (2025) Investigated the representation of AI ethics in the media discourse, revealed that moral concerns are central for public discussion about AI. The discovery is supported by the research of Kochetova (2023), who examined the cultural-linguistic illustration of AI in the English media, given that words such as "morality" and "Chabot" often coexist, underlines the intersection of technology with social values. Historical linguistic analysis complements these modern approaches by detecting the origin and spread of AI words. Brill (1993) documentation of the beginning of the term "Artificial Intelligence" around 1956, which marks the formal beginning of the region. Later words such as "nerve network" and "machine learning" have appeared in the 1960s and have since developed with scientific progress, as noted by Zhou (2023) and Luo & Lee (2022). Recent studies have also identified new linguistic incidents along with AI's progress. Kodirova (2025) documentation of the emergence of novel AI-related vocabulary, such as "Prompt Engineering" and "AI-Janit Material", reflects the effect of large language models such as GPT. Abdul Malik et al. (2022) Ana analyzed the discourse around the AI hallucinations, highlighting important approaches to the boundaries and challenges of AI technologies. Overall, literature indicates that the language of AI is versatile, constantly develops, and is closely associated with technical, social and moral development. This body of research further provides a fundamental understanding for corpus-based studies, which aims to track dynamic vocabulary and discourse pattern.

3.Methodology

3.1Data CollectionTo detect the effect of artificial intelligence on the development of English vocabulary, this study depends on a corpus-based functioning, which over time provides empirical evidence by analyzing the use of real language. The data was collected from several corpora which reflects different time periods and references to the use of English language. The primary source of data was the core of contemporary American English (Coca), which has been chosen since 1990 for widespread coverage of written and spoken texts. Coca provides a balanced perspective on newspapers, magazines, narratives, academic magazines and speaking conversations, which have developed vocabulary in various domains. Additionally, the study included data from news on the web (now) corpus, which provides up-to-date language samples from online news sources in various English speaking countries. This corpus was especially useful in tracing recent bounce in AI-related words and manifestations, especially after the emergence of large language models and AI-operated devices. To ensure a historical comparison, the Google Books Ngram viewer was also used to track the frequency of specific AI-related words from the beginning of the 20th century to the present time. This device helped identify wide historical trends and sequential integration of AI vocabulary in mainstream language. The keyword lists, collages, and frequency trends were extracted from these corpora using corpus analysis devices such as Antaconnac and Lancsbox. These devices allowed the researcher to track the emergence, frequency and relevant use of AI-related terminology (e.g. machine learning, algorithm, GPT, Chatbot etc.) and compared their growth over time.

The selection of these Corpora ensured that the data was representative of authentic, diverse and real -world English. By focusing on both historical and contemporary sources, the purpose of the study is to provide a comprehensive picture of how artificial intelligence has contributed to the dynamic development of English vocabulary. **Table 1. The Outline of AI-Generated Items**

Item No.	AI Term / Phrase	First Recorded Use	Corpus Source	Frequency (per million)	Context of Use
1	Machine Learning	1959	Google Books / COCA	12.5	Academic, Technology News
2	Neural Network	1960s	COCA / NOW	10.8	Scientific Journals, AI Research
3	Artificial Intelligence	1956	Google Books / COCA	15.3	General Media, Academia
4	Chatbot	1990s	NOW / Web Corpus	7.1	Customer Service, Social Media
5	GPT (Generative Pre-trained Transformer)	2018	NOW / News Articles	3.9	Technology Blogs, AI Discourse
6	Deep Learning	2006	COCA / Scientific Articles	9.4	Engineering, Data Science
7	Prompt Engineering	2022	NOW / Online Forums	2.3	AI Development, Tech Communities

Table 1 presents a structured observation of selected AI-related terminology items that emerged or gained prominence in the English language in recent decades. The table contains major words that were identified as a remarkable use pattern through corpus analysis, in both frequency and relevant variety. Each item has been listed with the approximate date of its earlier recorded use based on historical corpora such as Google Books and Coca to highlight the colonial development of AI vocabulary. For example, words such as "Artificial Intelligence" and "Machine Learning" are in the middle of the 20th century, reflecting their early concept in academic and scientific circles. In contrast, recent expressions such as "GPT" and "Prompt Engineering" portray the rapid vocabulary expansion powered by contemporary technological progress, especially after the rise of generic AI models. The corpus source column indicates the primary database from which the data of each word was extracted, ensuring the reliability of the use pattern. These sources - such as the corpus of contemporary American English (Coca), news on the web (now) corpus, and Google books - for representative samples of both formal and informal languages on different times. The frequency (per million) metric helps determine the relative prominence of each term, providing insight into which concepts are more widely used or embedded in public discourse. For example, the high frequency of "artificial intelligence" suggests its widespread recognition beyond technical areas. Finally, the reference to the use column highlights the specific domain or communication settings in which each word often appears. This qualitative dimension highlights the meaning and functional roles that play in modern English - from academic texts and scientific magazines to social media and customer service environment. Together, these elements explain how artificial intelligence has not only affected technological development, but has also shaped modern English lexicon in both form and function.

3.2Data Analysis The data collected from various corpora was analyzed using both quantitative and qualitative methods to examine the development of AI-related terminology in English. The purpose of the analysis is to track the emergence, frequency, corner behavior and semantic change of selected AI conditions over time. To begin, quantitative analysis focuses on the frequency trends of specific words in different time periods. Tools such as Antconc and Lancsbox were used to extract raw frequency data and to allow meaningful comparison incorporates

of various sizes to normalize. For example, increasing frequency of words like "deep learning", "neural network", and "Chabot" now reflects everyday news in the corpus and their growing use in public discourse. In addition to the frequency, the collecting analysis was organized to understand how these AI words are usually used in context. By examining frequent co-cum-words and phrases, the study identified the emerging patterns of meaning and use. For example, "AI" appeared with words such as "morality," "prejudice," "automation," and "innovation", which reflects its entry not only in technology but also in social and moral discussions. Qualitative analysis included in reading close to sample concurrent lines extracted from each corpus. This helped in capturing microscopic changes in meaning, tone and discourse functions. For example, while the previous uses of "artificial intelligence" were often limited to technical and scientific texts, recent use in media suggest a broad and sometimes more speculative or important tone. In addition, a comparative dichromic analysis was performed by examining the events of these conditions in historical corpora (e.g. Google books) and contemporary people (e.g. now, now, coca). This allowed the life cycle of the conditions to trace - from their initial emergence and adoption, mainstream acceptance and potentially cemented widening. Finally, the study paid special attention to neologisms and hybrid phrases such as "Prompt Engineering" and "AI-Janit Material", which were identified as signs of ongoing literal innovation. These items were evaluated on the basis of their relevant roles, meaning clarity and the rate of spread in various registers of English. Through this multi-dimensional approach, analysis provides a comprehensive approach about how artificial intelligence has actively shaped the development and expansion of modern English terminology.

Table 2. Frequency of Selected AI-Related Terms in COCA (1990–2020)

Term	1990s	2000s	2010s	2020–2023	Trend
Artificial Intelligence	112	258	624	1,095	Steady increase
Machine Learning	14	67	345	920	Sharp rise
Neural Network	48	123	270	415	Gradual growth
Deep Learning	0	5	122	650	Recent surge
Chabot	0	2	50	310	Rapid emergence

Table 2 reflects temporary development in the frequency of various AI-related words in about three and a half decades, reflecting the development and expansion of the region in the contemporary English discourse. A clear stable increase in the frequency of the term "Artificial Intelligence" has been observed, which increases from 112 incidents in the 1990s to more than 1,095 incidents during 2020-2023, indicating the growing public and educational awareness and interest in this domain. The term "machine learning" reflects an even faster growth, from only 14 times in the 1990s to 920 times in the most recent period, AI Research shows a major change towards machine learning techniques that ranks as the foundation of innovation and application This trend supports findings from studies such as Mikelionienė and Motiejūnienė (2021), who confirmed through corpus analysis that terms like "Machine Learning" and "Deep Learning" became more frequent after the mid-2000s, driven by big data revolutions and advances in computing capabilities. It also aligns with Udaya and Reddy (2024), whose work showed that modern corpora clearly reflect this rise in technical terms, influencing language education and curriculum development related to AI. Regarding the term "Deep Learning," it appeared relatively late in the 1990s but has experienced rapid growth over the last decade, reaching nearly 650 occurrences, indicating the widespread adoption of deep neural network techniques in practical applications. This finding corresponds with Zhou and Gao (2025), who noted that deep AI technologies lead contemporary scientific research and impact the evolution of related vocabulary. Finally, the term "Chatbot" emerges as a relatively new expression with noticeable gradual growth, emphasizing the rise of a new application type focused on human-machine interaction. This phenomenon was also observed in Kodirova (2023), who analyzed the cultural linguistic representation of AI in English media and noted increased usage of such application-oriented terms.

Table 3. Most Frequent Collocates with “Artificial Intelligence” in COCA

Collocate	Frequency	MI Score (Mutual Information)	Context of Use
System	89	5.8	Technical documents
Ethics	74	6.1	Academic, media
Technology	67	5.5	News, general discourse
Robot	59	5.3	Popular science, fiction
Human	48	4.9	Philosophical, ethical debates

This table presents words that frequently co-occur with the term "Artificial Intelligence," measuring their statistical association strength, thereby revealing the core contexts in which this term is used in modern English texts. The word "system" ranks first with 89 occurrences, indicating a strong link between AI and complex computer systems that integrate AI technologies in their operations. Following words such as "ethics" and "technology" emphasize the significance of ethical and technical dimensions in public discourse around AI, consistent with contemporary studies like Ormanova et al. (2025), which focused on the cultural linguistic aspects of these terms in media and academic debates, highlighting AI ethics as a central topic. The presence of words like "robot" and "human" reflects the close intertwining of technology and humanity, illustrating discussions about how AI integrates with daily human life or impacts it. This supports observations by Yunjiu et al. (2022), who compared AI-designed vocabulary tests with human ones, revealing how technological evolution influences language and context. Thus, the table confirms that AI is not discussed merely as a technical subject but also encompasses broad social, ethical, and human dimensions, as emphasized by Kochetova (2023) in her linguistic media analysis. Table 4. First Appearance of Terms in Google Books Ngram Viewer

AI Term	First Notable Appearance	Source Genre	Observation
Artificial Intelligence	1956	Academic texts	Coined at Dartmouth Conference
Machine Learning	1960s	Computer science papers	Gradual academic integration
Neural Network	1960s	Cognitive psychology	Early interdisciplinary usage
Deep Learning	~2006	AI research publications	Linked to neural net breakthroughs
Prompt Engineering	~2022	Technical blogs, forums	Directly related to LLMs (e.g., GPT-3, 4)

This table tracks the initial notable historical appearances of AI-related terms across literary and technical sources, with comments on the nature and context of these sources. The term "Artificial Intelligence" itself dates back to 1956, aligning with pioneering conferences that launched the field, as documented in Brill (1993), who recorded the linguistic origins of the field. Terms like "Machine Learning" and "Neural Network" emerged in the 1960s, reflecting early academic research combining computer science with cognitive psychology. This corresponds with Zhou (2023), who discussed building multilingual corpora and machine translation using dynamic models. The term "Deep Learning," appearing around 2006, marks a new wave of research leveraging advanced data processing techniques, leading to its rapid spread, as noted by Luo and Li (2022) in their study on machine translation. The recent emergence of "Prompt Engineering" in 2022 reflects a very recent development linked to large language models like GPT, highlighted in modern studies on AI applications in language and education, such as Szudarski (2025) in his corpus and educational literature review. This table demonstrates how the historical emergence of terms mirrors rapid technological and applied advances in AI, underscoring the importance of tracking lexical shifts in linguistic research.

Table 5. Contextual Use of “Chatbot” in NOW Corpus (2017–2023)

Year	Frequency (per million)	Typical Context	Sentiment Orientation
2017	1.2	Customer service, tech blogs	Neutral/positive
2019	3.5	Healthcare, education	Positive
2021	6.8	Ethics, misinformation, automation	Mixed
2023	9.3	AI assistants, LLMs (e.g., ChatGPT)	Varied (excitement/caution)

This table monitors changes in frequency and usage of the term "Chatbot" in recent years, highlighting usage contexts and associated sentiment trends. A strong increase is evident, from 1.2 occurrences per million words in 2017 to 9.3 in 2023, indicating widespread adoption of this technological application across diverse fields. Initially the references centered around customer service and technology have expanded to include healthcare and education, which indicates intelligent technologies in various life domains. The trends of emotion

reflect enthusiasm to caution, social and educational reactions in the form of chatbot benefits and concerns such as privacy or misinformation. It is aligned with Abdul Malik et al. (2022), whose corpus analysis of AI metaphors showed a series of positive and negative perceptions. Recent studies like Zhou and Gao (2025), also highlight the role of chatbot in increasing human-machine interaction and need to understand the use of language in this domain to ensure effective communication. Table 6. Neologisms and Emerging AI Terms (2020–2024)

Term	Type	Context	Corpus Evidence	Semantic Notes
Prompt Engineering	Compound Noun	Tech blogs, developer forums	NOW, Web Forums	Refers to crafting inputs for LLMs
AI-generated Content	Adjective Phrase	Media, education, journalism	COCA, NOW	Linked to content created by generative AI
Hallucination (AI)	Semantic Shift	AI criticism, NLP papers	NOW, ArXiv	AI producing incorrect/unreal information
Responsible AI	Phrase	Policy, ethics discussions	COCA, Google Books	Related to governance and fairness issues

This table reflects the presence of new terminology and emerging conditions in AI in recent years, which reflect linguistic mobility with technological progress. Words such as "Prompt Engineering" and "AI-Generated Content" give examples of large language models and linguistic changes with advanced AI applications. These observations coincide with Kodirova (2025), who documented the influence of automation and AI on technical and practical terminology, emphasizing new terms reflecting fundamental regional changes. In addition, concepts such as "AI hallucinations" have emerged as important words, which face challenges in the AI language model to generate misinformation, which highlights the semi- and linguistic dimensions of the area noted by Abdul Malik et al. (2022). The table confirms that the AI-related language is not only technical, but extends to include moral, technical and important words, requires updates for educational and research course, as stressed by Szudarski (2025) and Uday and Reddy (2024).

Conclusion

The findings obtained from the analysis of linguistic data in many databases clearly demonstrate the rapid growth and development of the field of artificial intelligence, not only from a technical point of view, but also from linguistic, social and cultural angles. It is clear that AI-related words like "Artificial Intelligence," "Machine Learning," and "Deep Learning" have seen a significant increase in use in recent decades, reflecting major technical changes and showing increasing interest from both researchers and general public. Additionally, moral and human dimensions of discussions around these techniques have been exposed through related words like "morality" and "chatbot". In addition, the emergence of new and developed terminology in recent years reflects the ongoing development and linguistic flexibility, emphasizes the need to include these changes in the educational course and incorporates research equipment for a comprehensive understanding of this rapidly growing area. In conclusion, this analysis indicates that studying the development of AI-related language provides valuable insight about how scientists and public communities interact with this domain. It also underlines the importance of continuous monitoring of linguistic and technical changes to ensure a permanent and responsible knowledge investment in the future of artificial intelligence.

References

- Abdul Malik, N., Ya Shak, M. S., Mohamad, F., & Joharry, S. A. (2022). Corpus-based studies of metaphor: An overview. *Arab World English Journal*, 13(2), 512-528.
- Brill, E. D. (1993). *A corpus-based approach to language learning*. University of Pennsylvania.
- Gabani, K., Solorio, T., Liu, Y., Hassanali, K. N., & Dollaghan, C. A. (2011). Exploring a corpus-based approach for detecting language impairment in monolingual English-speaking children. *Artificial intelligence in medicine*, 53(3), 161-170.

- Kochetova, L. A. (2023). Lingiocultural specifics of artificial intelligence representation in the English language media discourse: Corpus-based approach. *ВестникВолгоградского государственногоуниверситета. Серия 2: Языкознание*, 22(5), 6-18 [Bulletin of Volgograd State University. Series 2: Linguistics], 22(5), 6-18.
- Kodirova, D. S. (2005). Lexical borrowings and neologisms in the mining industry: The impact of automation and artificial intelligence. *Академическиеисследования в современной науке* [Academic Research in Modern Science], 4(28), 47-54.
- Li, D. (2024). The Existence and Development of Variants in English Writing Teaching in International Corpus based on Time Series Data Analysis. *Scalable Computing: Practice and Experience*, 25(6), 5585-5592.
- Liu, Y., & Liu, M. (2019, November). Research on college English teaching under the background of artificial intelligence. In *Journal of Physics: Conference Series* (Vol. 1345, No. 4, p. 042064). IOP Publishing.
- Luo, J., & Li, D. (2022). Universals in machine translation? A corpus-based study of Chinese-English translations by WeChat Translate. *International Journal of Corpus Linguistics*, 27(1), 31-58.
- Men, H. (2017). Vocabulary increase and collocation learning: *A corpus-based cross-sectional study of Chinese learners of English*. Springer.
- Mikelionienė, J., & Motiejūnienė, J. (2021). Corpus-based analysis of semi-automatically extracted artificial intelligence-related terminology. *Journal of language and cultural education.*, 9(1), 30-38.
- Ormanova, A. B., Anafinova, M. L., Ospanova, D. Z., & Tlesheva, Z. K. (2025). A Corpus-Based Approach in Vocabulary Research: Defining the Word of the Year 2023 in Kazakh. *Theory and Practice in Language Studies*, 15(3), 677-687.
- Szudarski, P. (2025). Vocabulary, corpus and language teaching. A machine-generated literature overview.
- Udaya, M., & Reddy, C. R. (2024). Corpus-Based Vocabulary Studies. In *Vocabulary, Corpus and Language Teaching: A Machine-Generated Literature Overview* (pp. 119-143). Cham: Springer Nature Switzerland.
- Yunjiu, L., Wei, W., & Zheng, Y. (2022). Artificial intelligence-generated and human expert-designed vocabulary tests: A comparative study. *Sage Open*, 12(1), 21582440221082130.
- Zhou, L. (2023). English-Chinese corpus collection and artificial intelligence translation based on dynamic clustering model. *International Journal of Science and Engineering Applications*, 148(58), 63-69.
- Zhou, W., & Gao, B. (2025). Construction and application of English-Chinese multimodal emotional corpus based on artificial intelligence. *International Journal of Human-Computer Interaction*, 41(3), 82-93.