Scientific landscape on opportunities and challenges of large language models and natural language processing

Rachel Edita O. Roxas, Reginald Neil C. Recario
Institute of Computer Science, College of Arts and Sciences, University of the Philippines Los Baños, Los Baños, Philippines

Article Info

Article history:

Received Feb 7, 2024
Revised Jun 2, 2024
Accepted Jun 5, 2024

Keywords:

Artificial intelligence
Bibliometrics
Large language models
Natural language processing
Opportunities and challenges

ABSTRACT

This paper conducted a systematic review of Scopus-indexed publications on large language models (LLMs) and natural language processing (NLP) extracted in October 2023 to address the dearth of literature on their opportunities and challenges. Through bibliometric analysis, from the 1,600 relevant documents, the study explored research productivity, revealing both opportunities and challenges spanning research and real-world applications in education, medicine, and health care, citations, and keyword co-occurrence networks. Results highlighted distribution patterns and dominant players like Google LLC and Stanford University. Opportunities such as technological development in generative artificial intelligence (AI), were contrasted with challenges such as biases and ethical concerns. The intellectual structure analysis revealed prominent application areas in health and education and also emphasized issues such as AI divide and human-AI partnership. Improvement on the technology performance of LLM and NLP remains to be a challenge. Recommendations include further exploration of open research problems and bibliometric studies using other research databases given the research bias towards Scopus-indexed English publications.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Rachel Edita O. Roxas
Institute of Computer Science, College of Arts and Sciences, University of the Philippines Los Baños
College, Laguna, Philippines
Email: roroxas2@up.edu.ph

1. INTRODUCTION

The field of artificial intelligence (AI) is increasingly permeating various sectors of society such as education [1], [2], health [3], [4], and finance and accounting [5]-[7] among others. The evident improvements in computer processors and the availability of libraries and methodologies that simplify the use of AI models and techniques help accelerate its popularity and reinforce the opportunities available. AI is an enabler in accomplishing 134 out of the 169 targets, across the 17 sustainable development goals [8], while AI has been shown to also inhibit 59 targets of the sustainable development goals [8]. Such cases exemplify how AI affects and influences the various aspects of our lives.

Large language models' (LLMs) role in reshaping the AI landscape cannot be overstated. These models, pre-trained on large numbers of data, have redefined the boundaries of what machines can achieve. LLMs, such as OpenAI's GPT-3, are no longer limited to text generation but can now perform various tasks from language translation to content creation.

Given the expansive applicability of LLMs and the consequent exponential increase in research and publications, review articles have been published to provide scientific landscapes for specific domains only, such as in the health sector [9], education [10], [11], medicine [12], legal informatics [13], and robotics [14]. Thus, this study will consider all and across domains of LLMs. It is also the case the reviews have been conducted using specific research databases such as the Web of Science (WoS) [15], while we use the Scopus database.

Against this backdrop, review publications have also highlighted the opportunities and challenges that are brought about by the utilization of LLMs in various fields [11], [16]-[19]. Opportunities include increased efficiency of delivery of services due to the automation of tedious and repetitive tasks, and descriptive and predictive analytics, to name a few. On the other hand, the use of LLMs also comes with challenges [13], [18], such as potential bias, misinformation, plagiarism, and other ethical concerns.

In the study of LLMs, natural language processing (NLP), which is the study of the computational aspects of natural languages, was identified by Fan *et al.* [20] as an integral component of LLM research. Thus, a systematic review of NLP publications was conducted by Krasadakis, Sakkopoulos, and Verykios [13], but it focused on NLP challenges only and was limited to the domain of legal informatics and low-resource languages. The review on LLMs by Fan *et al.* [20] particularly used publications that are listed on the WoS; however, the use of other research databases such as Scopus is an open area of research, which is the focus of this study.

Thus, to address the gap in existing literature, we conduct a systematic review of Scopus-indexed publications on the use of LLM and NLP across various domains and languages, but focusing on opportunities and challenges, which has not been addressed by earlier work. Through this study, we identify the scientific landscape of existing work, as well as open research problems in the LLM and NLP research field through a focus on opportunities and challenges. Our methodology is based on the bibliometric approach of Pritchard [21], and our data analysis and visualization are anchored on the work of Hallinger and Kovačevic [22] as our theoretical framework.

Bibliometric analysis is "a scientific computer-assisted review methodology that can identify core research or authors, as well as their relationship, by covering all the publications related to a given topic or field" as cited by Han *et al.* [23]. A bibliometric approach, as traced back to Pritchard [21], "employs mathematical and statistical methods in bibliometric to determine and analyze the growth and trend of a particular research theme" (p. 349), and many bibliometric studies have been conducted ever since, across various domains. Research databases such as Google Scholar [24], WoS [25], PubMed [9], and Scopus [26], have been used to conduct this type of quantitative approach [26], [27] and in various fields such as health policy and management [28], education [11], nursing [29], AI [30], agriculture [31], and educational research [32].

We describe the scientific landscape of the opportunities and challenges of the use of LLMs and NLP. We anchor our study on Hallinger and Kovačević [22] which presents the components of a bibliometric study, which are size, time, space, and composition. Size refers to both the quantity and quality of publications in the area which are extracted from the research database (which is presented in Table 1). We did not limit the time (the date of publication) or the space (the geographical distribution of the publications); all possible years of publication and all possible countries/territories of origin were included. The composition (or what is called intellectual structure) of the publications [23] includes visualizations using the Scopus research database function, Biblioshiny [33], and VOSviewer [34], and the identification of the application areas, nomenclatures, and open research areas.

The primary objective of this study is to investigate the scientific landscape of opportunities and challenges in LLM and NLP using the bibliometric approach, and to provide a holistic analysis of the field, revealing its overall structure and dynamics.

Specifically, the research aims:

- to analyze the trends and patterns of Scopus-indexed publications on LLMs and NLP opportunities side-by-side challenges by country/territory, subject area, affiliation, funding sponsor, citation, and keyword co-occurrence; and
- to evaluate the patterns of Scopus-indexed publications on LLMs and NLP opportunities and challenges by its intellectual structure focusing on application areas, nomenclatures, and open research areas.

2. METHOD

This study employed a systematic literature review using the bibliometric approach for critical discussion, comparison, and interpretation of the analyses. This study has been anchored on the work of Hallinger and Kovačević [22] as our theoretical framework, with data collection from the Scopus research database, based on the preferred reporting items for systematic reviews and meta-analysis (PRISMA) procedure [35]. Then we performed data analysis and visualizations based on the work of Scopus, Biblioshiny [33], and VOSviewer [34], including intellectual structure or composition of the documents [23] using keyword co-occurrence networks [36], and the identification of application areas, nomenclatures, and open research areas. Details are presented in the succeeding discussions.

2.1. Collection of publication data

The PRISMA procedure was adopted [35]. The PRISMA procedure has three stages, namely, identification, screening, and included. At the identification phase, documents were retrieved during October 2023 with boolean queries in the advanced search function of Scopus (as shown in Table 1). A search using TITLE-ABS-KEY ("large language model" OR LLM) yielded 1,600 documents after the screening and exclusion of non-English documents. Table 1 presents the search functions conducted to identify the documents on the areas of LLM, NLP, opportunities and challenges, and the resulting number of documents as identified by the search functions. The criteria for inclusion of documents in this study are those published and indexed on the Scopus research database, published on or before October 2023, and those that are written in English.

Then we extracted from Scopus the metadata of these documents to comma-separated value (CSV) files with the corresponding number of documents. Each row in the CSV file contains the metadata of one document that is considered in this study, which includes the names of the authors, the title of the document, the abstract, the affiliation, and other pertinent information about the publication.

Table 1. Search functions and numbers of extracted documents

Search function	Number of extracted documents
TITLE-ABS-KEY ("large language model" OR LLM)	1,600
TITLE-ABS-KEY (("large language model" OR LLM) AND ("natural language processing" OR NLP) AND (promise OR opportunity OR future OR blessing))	96
TITLE-ABS-KEY (("large language model" OR LLM) AND ("natural language processing" OR NLP) (challenges OR problems OR concern OR curse))	149
TITLE-ABS-KEY (("large language model" OR LLM) AND ("natural language processing" OR NLP) AND (promise OR opportunity OR future OR blessing)) AND (challenges OR problems OR concern OR curse))	44

2.2. Data visualizations

Using the Scopus function analyze-search-results, the graphs for documents by affiliation, country or territory, affiliation, subject area, and citation were generated and correspondingly exported into a storage device. Further analyses and visualizations were performed using Biblioshiny [33] and VOSviewer [34] using the CSV files exported from Scopus of the metadata of the documents considered in this study (described in the previous section). Biblioshiny [33] was used to generate the main information, and VOSviewer [34] was used to generate the keyword co-occurrence network.

Keyword co-occurrence has been widely used in mapping the knowledge structure of publications in particular fields of study [36] and performs clustering, where keywords are automatically grouped into clusters which are represented by different colors if keywords co-occur in documents. The particular settings for each generation of the network have been included in the next section.

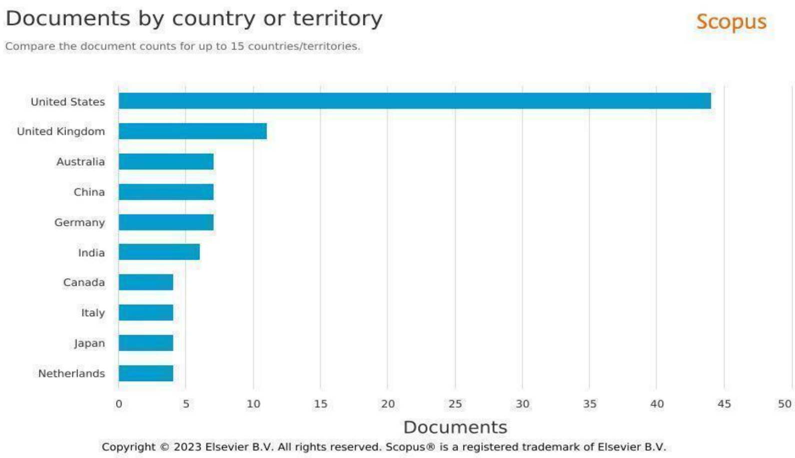
3. RESULTS AND DISCUSSION

This section contains the analyses performed and some inferences in the area of LLMs, and LLMs and NLP opportunities and challenges. After searching for 1,600 Scopus-indexed publications, we performed analyses of the 96 LLM and NLP opportunities side-by-side the 149 LLM and NLP challenges documents by country/territory, subject area, affiliation, funding sponsor, citation, and keyword co-occurrences. Thereafter, we assessed the trends and patterns of the 44 Scopus-indexed publications of the LLMs and NLP opportunities and challenges by its composition (or what is called intellectual structure according to Hallinger and Kovačević [22]) in terms of various application areas, nomenclatures, and open research areas.

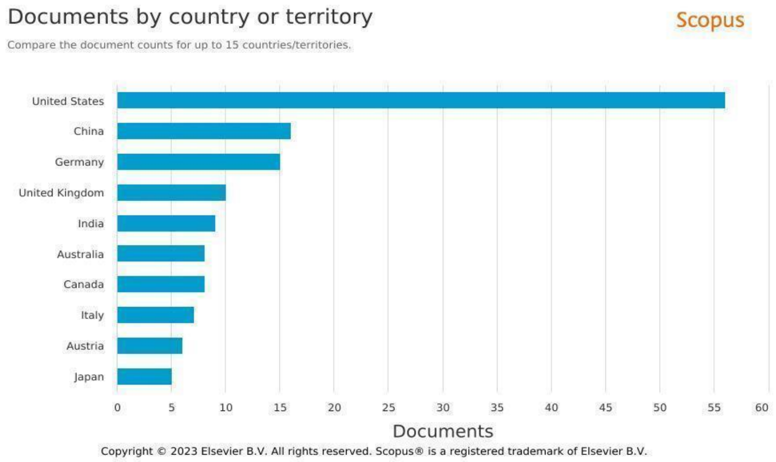
3.1. LLM and NLP: opportunities side-by-side challenges

This section discusses the trends and patterns of the 96 LLM and NLP opportunities side-by-side the 149 on challenges (as listed in Table 1) by considering the most documents by opportunities and challenges respectively by country/territory as shown in Figures 1(a) and 1(b), subject area as shown in Figures 2(a) and 2(b), affiliation as shown in Figures 3(a) and 3(b), funding sponsor as shown in Figures 4(a) and 4(b), top 10 cited documents, and keyword co-occurrence networks as shown in Figures 5(a) and 5(b).

We further examined the publications on LLM and NLP by looking at the distribution of documents by country or territory (Figure 1), focusing on opportunities and challenges, as shown in Figures 1(a) and 1(b), respectively. The United States leads in both the opportunities and challenges for LLM and NLP, followed by the United Kingdom for opportunities, and China in the challenges. It can be noted that several countries or territories in the top 10 are both present for the opportunities and challenges. Despite the benefits that AI provides to society, it also introduces an AI-divide, where low-income countries, societies, organizations, and/or individuals who have been alienated in the digital-divide due to the limited accessibility to digital devices and tools, are further alienated in the advent of AI [37]. Review publications on LLM and NLP fail to mention about the concept of AI-divide such as in [12], [13].



(a)



(b)

Figure 1. Documents by countries/territories of LLM and NLP (a) opportunities and (b) challenges

We further examined the publications on LLM and NLP by looking at the distribution of documents by subject area (Figure 2), focusing on opportunities and challenges, as shown in Figures 2(a) and 2(b), respectively. For both opportunities and challenges, Computer Science dominates the subject areas with 37.7% and 42.8% respectively, showing that publications on technology development recognize that there are both benefits and problems associated with the advancement of LLM and NLP; however, there is a higher percentage associated with the challenges. It is followed by Social Sciences with 11.9% for opportunities and 11.8% for challenges, with almost the same percentages, revealing that these publications focus on societal aspects of the introduction of these technologies. The remaining subject areas show that these technologies are present in a variety of fields, including medicine, arts and humanities, and business, but also may present both opportunities and challenges.

We looked at the distribution of documents by top 10 affiliations (Figure 3), focusing on opportunities and challenges, as shown in Figures 3(a) and 3(b), respectively. In the top 10 institutions of both LLM and NLP opportunities and challenges, the top 1 and top 2 are occupied by Google LLC and Stanford University, respectively. Both the top 10 on opportunities and challenges are dominated by Universities, but with a greater involvement of industry-related institutions in challenges than in opportunities (1:3 with only Google LLC in opportunities, in contrast to Google LLC, IBM research, and

Amazon.com, in challenges), and greater involvement of medicine-related institutions in opportunities (School of Clinical Medicine, and Mayo Clinic) with none in challenges. From these figures, academe and industry involvements are evident.

We examined the distribution of documents by funding sponsor (Figure 4), focusing on opportunities and challenges, as shown in Figures 4(a) and 4(b), respectively. A health-related funding sponsor, the National Institutes of Health, highlighted opportunities. Various private and government institutions focusing on health, sciences, military, education, and culture have funded and shown interest in LLM and NLP for their various fields of study, highlighting both the opportunities and challenges encountered.

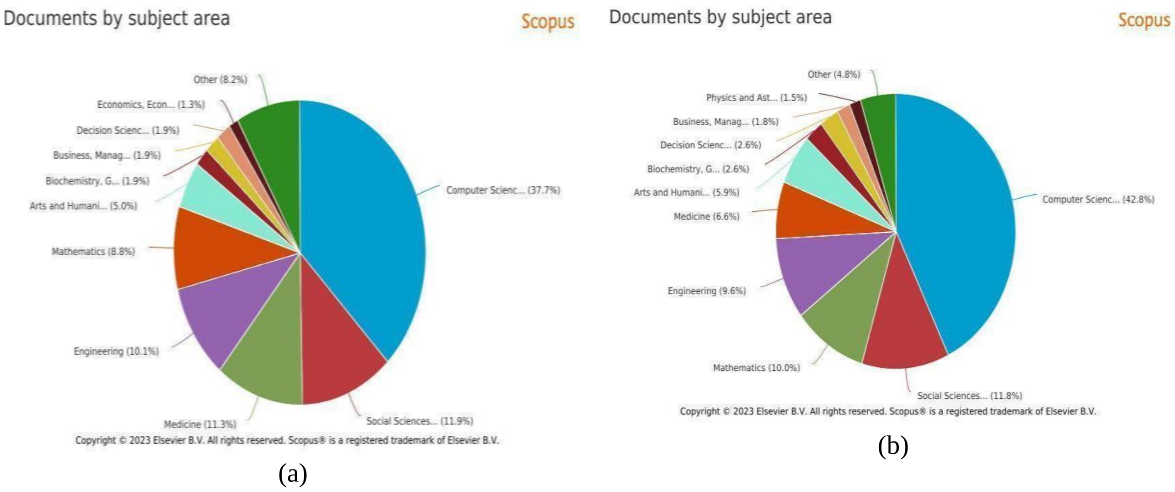


Figure 2. Documents by subject area of LLM and NLP (a) opportunities and (b) challenges

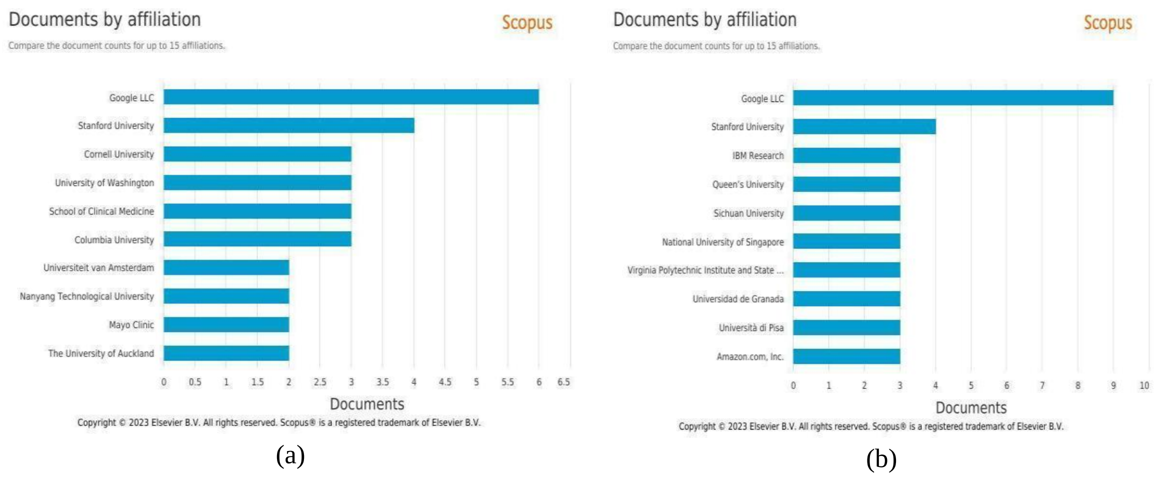
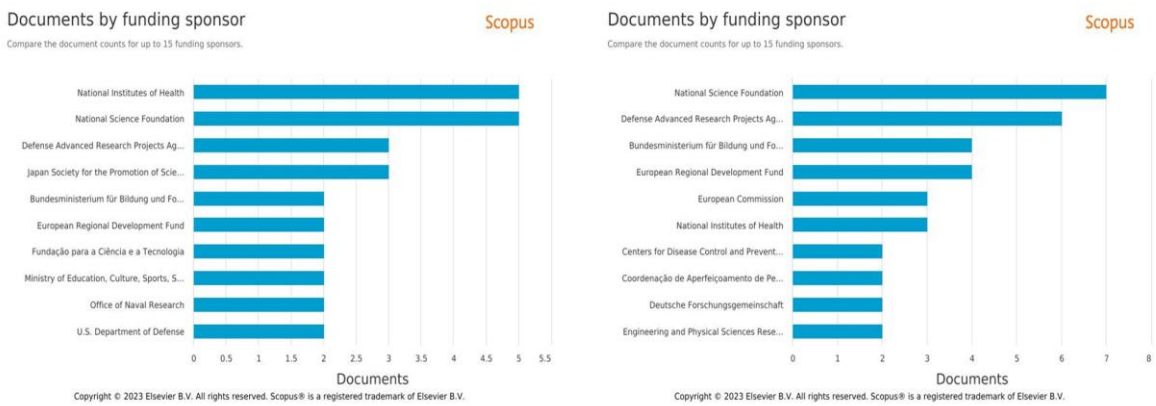


Figure 3. Documents by affiliation on LLM and NLP (a) opportunities and (b) challenges

The top 10 cited documents among the 96 LLM and NLP opportunities side-by-side the 149 LLM and NLP challenges were analyzed. The top 10 cited documents of the publications about LLM and NLP opportunities obtained a total of 451 citations with the top 1 cited document [38] on detection and correction of spelling errors using LLMs having 102 citations. Common topics include human-AI interaction and collaboration (3 out of 10 or 30%), ChatGPT (3 out of 10 or 30%), and GPT-3 (2 out of 10 or 20%). Human-AI interaction and collaboration are termed as co-author [39] and co-writing [40] between human and AI. These are new findings in contrast to bibliometric studies on LLM publications [10], [20] where these were not mentioned in their findings, except for the review paper on LLMs on human-robot interaction for improved performance, where the human provides intent to the robot and human actions to the environment, while the robot provides feedback to the human and robot actions to the environment [14]. Domains include biotechnology in protein sequencing [41], spelling error detection and correction [38], clinical practice [42], education [43], public health [44], and code explanations [45]. Meanwhile, the top 10 cited documents among the 149 LLM and NLP challenges obtained a total of 266 citations with the top 1 document on code explanations [44] having 22 citations. These documents mainly discussed GPT (1 out of 10 or 10%) and ChatGPT (4 out of 10 or 40%), and focused on challenges associated with its use in the following domains: public health [44] and clinical practice [42], education [43], [46], [47], and coding [45]. For education, the specific areas include medical education [46], private learners [47], and standard admission tests [43]. Challenges include fake news [48], code explanations [45], and detoxifying LLMs [49].



(a) (b)

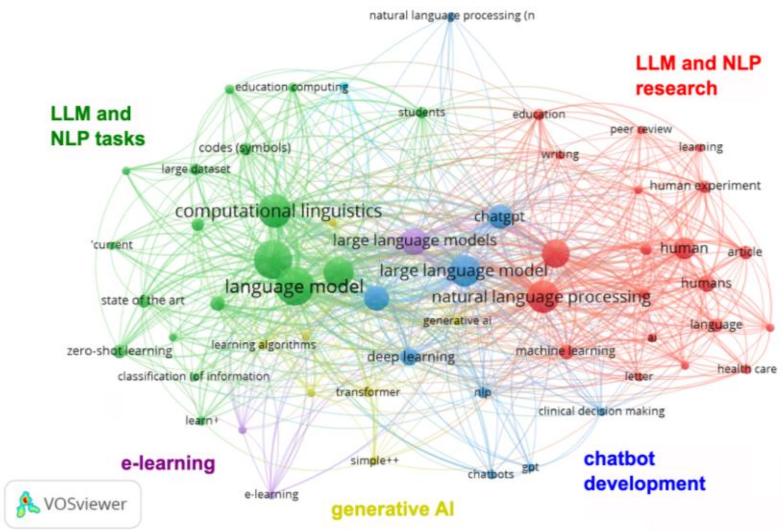
Figure 4. Documents by funding sponsor of LLM and NLP (a) opportunities and (b) challenges

The top 10 cited documents among the 96 LLM and NLP opportunities and the 149 LLM and NLP challenges have common documents [42]-[44], [50], and common domains, such as education, public health, and code explanations. Code explanations come under a field called explainable artificial intelligence (XAI) that hopes to transition from "black box" methods of LLMs into new approaches that enables explanation and interpretability [51]. This is in support of the findings of existing systematic reviews on LLMs, when LLM-based systems, although providing correct answers to questions, outputs "nonsensical explanations" and "highly misleading" [9], simply mimic human behavior like "stochastic parrots" without full understanding of how the answers were derived [20] and does not allow human oversight on outputs of LLM systems, especially in sensitive domains such as healthcare and law [15]. In contrast, explain ability has not been mentioned in other LLM review publications [10]. Such common documents and domains imply that these domains present opportunities, and also pose as challenges, and vice versa. We can also infer that both opportunities and challenges of LLM and NLP overall influence the political economy of AI [50].

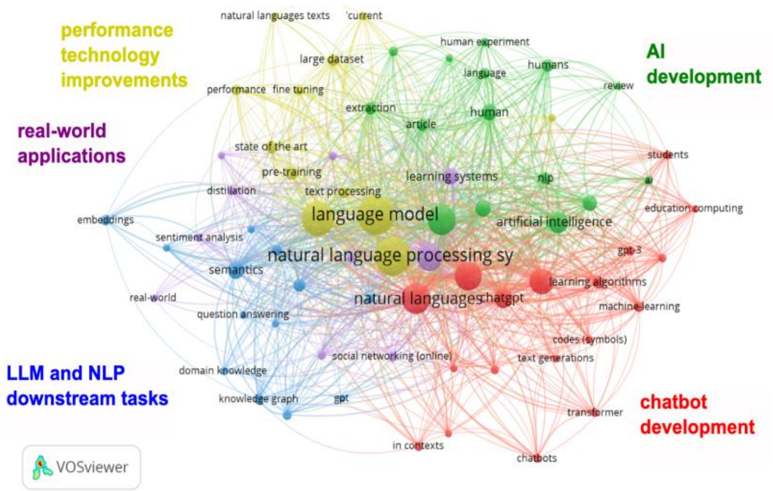
We present the keyword co-occurrence networks of publications on LLM and NLP (Figure 5), focusing on opportunities and challenges, as shown in Figures 5(a) and 5(b), respectively. Figure 5(a) presents the keyword co-occurrence network generated from 96 LLM NLP opportunities using all keywords, minimum number of occurrences of a keyword=3, of the 839 keywords, 58 meet the threshold, has 5 clusters (identified by different colors), which were manually labeled as follows: LLM and NLP research (red), LLM and NLP tasks (green), chatbot development (blue), generative AI (yellow), and e-learning (purple). The red cluster, which is on LLM and NLP research, includes keywords that are opportunities for various application areas such as education, health care, and medical research. The green cluster highlights the opportunities for the further development of LLM and NLP tasks such as natural language generation, classification, and text processing. The blue cluster is dominated by keywords associated with deep learning, specifically GPT, towards the development of chatbots. The yellow cluster focuses on generative AI using transformer architectures. Finally, the purple cluster highlights the LLM NLP opportunities in e-learning.

Figure 5(b) illustrates the keyword co-occurrence network, generated by VOSViewer, based on 149 LLM NLP challenges. This network encompasses all keywords, with a minimum occurrence threshold set at 4 out of the 1,176 total keywords. Among these, 68 keywords meet the threshold, revealing 5 distinct clusters represented by a different color and manually labeled as follows: chatbot development (red), AI development (green), LLM and NLP downstream tasks (blue); performance technology improvements (yellow); and real-world applications (purple). The red cluster, chatbot development, though generated as a cluster in the opportunity's documents, has also been seen as an LLM NLP challenge. It includes in-contexts, and applications to computing education. The green cluster highlights the challenge of AI-human relationship and partnership. The blue cluster includes LLM and NLP downstream tasks such as named entity recognition, text classification, and question answering systems. The yellow cluster focuses on data to improve performance of training and fine tuning. The last cluster, the purple cluster, highlights the real-world applicability of LLM NLP applications such as areas of finance through social media data and sentiment analysis. These clusters offer a nuanced understanding of the diverse challenges within the LLM NLP domain, encompassing technical intricacies and real-world implementation concerns.

From the keyword co-occurrence networks of LLMs and NLP opportunities as shown in Figure 5(a), and LLMs and NLP challenges in Figure 5(b), common clusters include technical areas of development on AI, Chatbots, and LLM and NLP tasks. Clusters were more specific, referring to generative AI in opportunities; on the other hand, for challenges, LLM and NLP downstream tasks, and technology performance improvements were included. These results show the dominance of the computer science subject area, as similarly reflected by Figures 2(a) and 2(b), for both publications on opportunities and challenges. What stands out in the findings is the cluster of publications in LLM and NLP challenges on AI-human relationship and partnership, which was not mentioned in other reviews on LLMs [9], [10], [12], but the proposed use of the feedback "thumbs down" button to improve the performance of LLM systems is a form of human-AI partnership towards better accuracy [12], [14].



(a)



(b)

Figure 5. Keyword co-occurrence network on (a) 96 LLM and NLP opportunities and (b) 149 LLM and NLP challenges

3.2. LLMs and NLP opportunities and challenges: intellectual structure

We analyzed the 44 documents on LLMs and NLP on both opportunities and challenges in terms of various application areas, nomenclatures, and open research areas. Those domains that stand out in the literature relate to health and education. In the area of health, LLM application areas include public health [44], health care [52], [53], mental health [54], dentistry [55], biomedical research [56], surgery [57], in vitro fertilization [53], pediatrics [58], and clinical practice [42]. In education, studies covered academia, in general [11], [59], [60], while some focused on higher education and research [1], engineering education [61], standardized admissions test [43], creative writing [62], language learning [63], and peer feedback [64].

Other application areas include the use of autonomous driving systems [65], law [66], finance and accounting [67], [68], cybersecurity [69], human resource management [70], remote sensing [71], railway design [72], sexuality studies [73], and programming [74]. It is interesting to note that although programming has been addressed as an LLM application area [74], the collaboration of human and AI towards the design and development of software-intensive systems has been addressed by Ahmad *et al*. [75], which has been found to improve bottomline in business applications [76].

While LLM applications have revolutionized various domains, there are still limitations that have been the focus of publications. Low-resource languages and domains perform poorly in LLM systems [77], necessitating the fine-tuning and contextualization of LLM systems for better performance across various domains, diverse languages, and different times (or temporal information). For instance, certain LLMs are augmented with data on specific languages to improve performance; an example is AlephBERT, which is an LLM for modern Hebrew [78]. This is in support of the findings and recommendation of existing systematic review on LLMs, that researchers should endeavor to focus on the research on LLM and NLP towards "specialized language models" but as applied to healthcare [9], and to education [11]. To add, ambiguity is a big issue in NLP, and temporal information allows resolution of word sense ambiguity [79]. LLM applications handle various modes of data, such as text, audio, image, and video, or collectively called multi-modal data [55], [80].

Prompt engineering is also one area that can either limit or expand the capabilities of the LLM application [81]-[84]. Other studies have proposed the use of AI to detect AI-generated texts, which could be a useful tool for various domains [59], [85]. This confirms the findings in a review paper on LLMs as applied to medicine and education [10], [11].

In terms of nomenclature, the 44 documents on LLMs and NLP opportunities and challenges used the following terms for "challenges" in various forms and degrees: challenges [55], [19], concerns [11], pitfalls [56], and threats [85], to name a few. The challenges mentioned in these documents include: fair use [11], infodemic or misinformation or fake information [44], [56], data privacy [55], [60], data quality [55], trust [19], security [66], bias [11], [55], [50], explain ability [58], replicability and transparency, beneficence considerations [60], and ethics [61]. It is interesting to note that explain ability is a LLM and NLP challenge that was not mentioned in the bibliometric study on LLM reviews [12], [20].

Other concerns that have been raised by literature on the role of AI research and development side of using LLMs on the political economy includes: environmental footprints, and future social effects [50] towards ethical and responsible development and deployment [54], such as the commodification of LLMs as evidenced by paid API access and usage, and the continuous divide between those with and without access to computing resources [50]. Although the AI divide becomes wider, LLM applications are also perceived as enablers towards inclusivity of particular sectors of society such as an e-mail writing assistant for adults with dyslexia [86].

In summary, LLM and NLP opportunities are also its challenges, and vice versa. Challenges present opportunities for new research; with new opportunities and the capability to advance society, there also come challenges with ethical and social ramifications. As rightly expressed by Chen and Lin [87], AI is a "double-edged sword"; as opportunities abound for the use of AI in our everyday life, challenges also abound.

4. CONCLUSIONS

In our work, we present the scientific landscape of the Scopus-indexed documents of LLMs and NLP and opportunities and challenges using the bibliometric approach. We analyzed the trends and patterns of Scopus-indexed publications on LLMs and NLP opportunities side-by-side challenges by country/territory, subject area, affiliation, funding sponsor, citation, and keyword co-occurrence; and evaluated the patterns of Scopus-indexed publications on LLMs and NLP opportunities and challenges by its intellectual structure focusing on application areas, nomenclatures, and open research areas. In the analysis on publications on LLMs and NLP opportunities side-by-side challenges by country/territory, the study confirms a widening AI-divide among countries, societies, organizations, and individuals, who have been alienated in the digital-divide due to limited accessibility to digital devices and tools. This study laid out this new research area in contrast to the bibliometric study on LLM publications [20].

In the subject area analysis in opportunities side-by-side challenges, we observe a dominance of the computer science subject area, which shows the technological development and consequent publication of technological advancement, and the social science subject area, which reveals the crucial discourse on societal issues and concerns within these technological advancements. In the affiliation analysis, it is observed that academe and industry involvement as affiliation of authors is a very important component in the advancement of research. In the funding sponsor analysis, publications on opportunities and challenges cut across various private and government institutions, focusing on various areas. In the citation analysis of the top 10 cited documents, common domains include those in education, public health, and code explanations, implying that these domains present opportunities and also pose challenges, and vice versa. In the keyword co-occurrence networks, common clusters include technical areas of development on AI, Chatbots, and LLM and NLP tasks, and the further improvement of technology performance as a very important challenge. Similar to the subject area analysis, the keyword co-occurrence results show the dominance of the computer science subject area for publications on opportunities and challenges.

We also recommend the exploration of open research problems as identified in this study. In the technology aspect, open problems include more effective approaches towards better performance of LLMs and NLP. On the social science side, topics include human-AI relationship, partnership, and collaboration, and topics covered by ethical AI such as explainable AI, inclusivity, and the widening AI-divide. We must come up with AI solutions in the detection of unethical uses of AI-generated outputs. This study has discovered these new research areas in contrast to other bibliometric studies on LLM publications. Both opportunities and challenges must be addressed in research. Note that this study is particularly biased towards Scopus-indexed publications, and those that are written in English. Thus, we recommend exploration of scientific landscapes using other research databases, such as Google Scholar and WoS, or a combination of these research databases. We also recognize that due to the specific focus of our study on opportunities and challenges, we have considered here a relatively small number of documents (46, 149, and 44), in contrast to other review papers that considered thousands of publications. Overall, publications on LLM and NLP opportunities and challenges reveal that opportunities present new challenges, while new challenges become opportunities, much like a double-edged sword.

REFERENCES

[1] T. Alqahtani *et al.*, "The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research," *Research in Social and Administrative Pharmacy*, vol. 19, no. 8, pp. 1236–1242, Aug. 2023, doi: 10.1016/j.sapharm.2023.05.016.

[2] L. Chen, P. Chen, and Z. Lin, "Artificial intelligence in education: a review," *IEEE Access*, vol. 8, pp. 75264–75278, 2020, doi: 10.1109/ACCESS.2020.2988510.

[3] U. Pagallo *et al.*, "The underuse of AI in the health sector: opportunity costs, success stories, risks and recommendations," *Health and Technology*, vol. 14, no. 1, pp. 1–14, Jan. 2024, doi: 10.1007/s12553-023-00806-7.

[4] J. Bajwa, U. Munir, A. Nori, and B. Williams, "Artificial intelligence in healthcare: transforming the practice of medicine," *Future Healthcare Journal*, vol. 8, no. 2, pp. e188–e194, Jul. 2021, doi: 10.7861/fhj.2021-0095.

[5] A. K. V. N. Biju, A. S. Thomas, and J. Thasneem, "Examining the research taxonomy of artificial intelligence, deep learning & machine learning in the financial sphere—a bibliometric analysis," *Quality and Quantity*, vol. 58, no. 1, pp. 849–878, Feb. 2024, doi: 10.1007/s11135-023-01673-0.

[6] P. Shakdwipee, K. Agarwal, H. Kunwar, and S. Singh, "Artificial intelligence in finance and accounting: opportunities and challenges," in *Lecture Notes in Networks and Systems*, vol. 765 LNNS, 2023, pp. 165–177.

[7] L. Cao, "AI in Finance: A Review," *SSRN Electronic Journal*, 2020, doi: 10.2139/ssrn.3647625.

[8] R. Vinuesa *et al.*, "The role of artificial intelligence in achieving the sustainable development goals," *Nature Communications*, vol. 11, no. 1, p. 233, Jan. 2020, doi: 10.1038/s41467-019-14108-y.

[9] J. Li, A. Dada, B. Puladi, J. Kleesiek, and J. Egger, "ChatGPT in healthcare: a taxonomy and systematic review," *Computer Methods and Programs in Biomedicine*, vol. 245, p. 108013, Mar. 2024, doi: 10.1016/j.cmpb.2024.108013.

[10] E. Tikhonova and L. Raitskaya, "ChatGPT: where is a silver lining? Exploring the realm of GPT and large language models," *Journal of Language and Education*, vol. 9, no. 3, pp. 5–11, Sep. 2023, doi: 10.17323/jle.2023.18119.

[11] J. G. Meyer *et al.*, "ChatGPT and large language models in academia: opportunities and challenges," *BioData Mining*, vol. 16, no. 1, p. 20, Jul. 2023, doi: 10.1186/s13040-023-00339-9.

[12] T. Dave, S. A. Athaluri, and S. Singh, "ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations," *Frontiers in Artificial Intelligence*, vol. 6, May 2023, doi: 10.3389/frai.2023.1169595.

[13] P. Krasadakis, E. Sakkopoulos, and V. S. Verykios, "A survey on challenges and advances in natural language processing with a focus on legal informatics and low-resource languages," *Electronics (Switzerland)*, vol. 13, no. 3, p. 648, Feb. 2024, doi: 10.3390/electronics13030648.

[14] C. Zhang, J. Chen, J. Li, Y. Peng, and Z. Mao, "Large language models for human–robot interaction: a review," *Biomimetic Intelligence and Robotics*, vol. 3, no. 4, p. 100131, Dec. 2023, doi: 10.1016/j.birob.2023.100131.

[15] D. M. Petroșanu, A. Pîrjan, and A. Tăbușcă, "Tracing the influence of large language models across the most impactful scientific works," *Electronics (Switzerland)*, vol. 12, no. 24, p. 4957, Dec. 2023, doi: 10.3390/electronics12244957.

[16] D. Myers *et al.*, "Foundation and large language models: fundamentals, challenges, opportunities, and social impacts," *Cluster Computing*, vol. 27, no. 1, pp. 1–26, Nov. 2024, doi: 10.1007/s10586-023-04203-7.

[17] B. Zsidai *et al.*, "A practical guide to the implementation of AI in orthopaedic research – part 1: opportunities in clinical application and overcoming existing challenges," *Journal of Experimental Orthopaedics*, vol. 10, no. 1, Jan. 2023, doi: 10.1186/s40634-023-00683-z.

[18] K. Denecke, R. May, and O. Rivera-Romero, "Transformer models in healthcare: a survey and thematic analysis of potentials, shortcomings and risks," *Journal of Medical Systems*, vol. 48, no. 1, p. 23, Feb. 2024, doi: 10.1007/s10916-024-02043-5.
[19] S. S. Sohail *et al.*, "Decoding ChatGPT: a taxonomy of existing research, current challenges, and possible future directions," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 8, p. 101675, Sep. 2023, doi: 10.1016/j.jksuci.2023.101675.
[20] L. Fan, L. Li, Z. Ma, S. Lee, H. Yu, and L. Hemphill, "A bibliometric review of large language models research from 2017 to 2023," *ACM Transactions on Intelligent Systems and Technology*, May 2024, doi: 10.1145/3664930.
[21] P. Alan, "Statistical bibliography or bibliometrics," *Journal of Documentation*, vol. 25, pp. 348–349, 1969, [Online]. Available: https://www.researchgate.net/publication/236031787_Statistical_Bibliography_or_Bibliometrics%0Apapers2://publication/uuid/C6D8F1CE-3B68-45E7-8FA9-5CCB49CB2AE8.
[22] P. Hallinger and J. Kovačević, "Mapping the intellectual lineage of educational management, administration and leadership, 1972–2020," *Educational Management Administration and Leadership*, vol. 50, no. 2, pp. 192–216, Mar. 2022, doi: 10.1177/17411432211006093.
[23] J. Han, H. J. Kang, M. Kim, and G. H. Kwon, "Mapping the intellectual structure of research on surgery with mixed reality: Bibliometric network analysis (2000–2019)," *Journal of Biomedical Informatics*, vol. 109, p. 103516, Sep. 2020, doi: 10.1016/j.jbi.2020.103516.
[24] R. K. Bannor and K. K. Arthur, "A systematic review and bibliometric analysis on agribusiness gaps in emerging markets," *Research in Globalization*, vol. 8, p. 100214, Jun. 2024, doi: 10.1016/j.resglo.2024.100214.
[25] Y. Rivero-Moreno *et al.*, "Robotic surgery across Latin America: a bibliometric analysis of research trends from 2009 to 2022," *Journal of Robotic Surgery*, vol. 18, no. 1, p. 123, Mar. 2024, doi: 10.1007/s11701-024-01865-2.
[26] A. Martín-Martín, E. Orduna-Malea, and E. D. López-Cózar, "Coverage of highly-cited documents in Google Scholar, Web of Science, and Scopus: a multidisciplinary comparison," *Scientometrics*, vol. 116, no. 3, pp. 2175–2188, Sep. 2018, doi: 10.1007/s11192-018-2820-9.
[27] K. Şahin and G. Candan, "Scientific productivity and cooperation in Turkic world: a bibliometric analysis," *Scientometrics*, vol. 115, no. 3, pp. 1199–1229, Jun. 2018, doi: 10.1007/s11192-018-2730-x.
[28] F. Fusco, M. Marsilio, and C. Guglielmetti, "Co-production in health policy and management: a comprehensive bibliometric review," *BMC Health Services Research*, vol. 20, no. 1, p. 504, Dec. 2020, doi: 10.1186/s12913-020-05241-2.
[29] F. G. V. Jabonete and R. E. O. Roxas, "Barriers to research utilization in nursing: a systematic review (2002–2021)," *SAGE Open Nursing*, vol. 8, p. 237796082210910, Jan. 2022, doi: 10.1177/23779608221091073.
[30] A. Tiwari, S. Bardhan, and V. Kumar, "A bibliographic study on artificial intelligence research: global panorama and indian appearance," *Library Herald*, vol. 60, no. 4, pp. 15–36, 2022, doi: 10.5958/0976-2469.2022.00036.7.
[31] D. D. Mühl and L. de Oliveira, "A bibliometric and thematic approach to agriculture 4.0," *Heliyon*, vol. 8, no. 5, p. e09369, May 2022, doi: 10.1016/j.heliyon.2022.e09369.
[32] P. Song and X. Wang, "A bibliometric analysis of worldwide educational artificial intelligence research development in recent twenty years," *Asia Pacific Education Review*, vol. 21, no. 3, pp. 473–486, Sep. 2020, doi: 10.1007/s12564-020-09640-2.
[33] M. Aria, C. Cuccurullo, L. D'aniello, M. Misuraca, and M. Spano, "Thematic analysis as a new culturomic tool: the social media coverage on COVID-19 Pandemic in Italy," *Sustainability (Switzerland)*, vol. 14, no. 6, p. 3643, Mar. 2022, doi: 10.3390/su14063643.
[34] N. J. van Eck and L. Waltman, "{VOSviewer} manual," *Leiden: Univeristeit Leiden*, 2013, [Online]. Available: http://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.1.pdf.
[35] M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *Systematic Reviews*, vol. 10, no. 1, p. 89, Dec. 2021, doi: 10.1186/s13643-021-01626-4.
[36] H. N. Su and P. C. Lee, "Mapping knowledge structure by keyword co-occurrence: A first look at journal papers in Technology Foresight," *Scientometrics*, vol. 85, no. 1, pp. 65–79, Oct. 2010, doi: 10.1007/s11192-010-0259-8.
[37] X. Wang *et al.*, "ChatGPT: promise and challenges for deployment in low- and middle-income countries," *The Lancet Regional Health - Western Pacific*, vol. 41, p. 100905, Dec. 2023, doi: 10.1016/j.lanwpc.2023.100905.
[38] M. Gamon *et al.*, "Using Contextual Speller Techniques and Language Modeling for ESL Error Correction," *IJCNLP 2008 - 3rd International Joint Conference on Natural Language Processing, Proceedings of the Conference*, vol. 1, pp. 449–456, 2008.
[39] M. Lee, P. Liang, and Q. Yang, "CoAuthor: designing a human-AI collaborative writing dataset for exploring language model capabilities," in *Conference on Human Factors in Computing Systems - Proceedings*, Apr. 2022, pp. 1–19, doi: 10.1145/3491102.3502030.
[40] A. Yuan, A. Coenen, E. Reif, and D. Ippolito, "Wordcraft: story writing with large language models," in *International Conference on Intelligent User Interfaces, Proceedings IUI*, Mar. 2022, pp. 841–852, doi: 10.1145/3490099.3511105.
[41] A. Madani *et al.*, "Large language models generate functional protein sequences across diverse families," *Nature Biotechnology*, vol. 41, no. 8, pp. 1099–1106, Aug. 2023, doi: 10.1038/s41587-022-01618-2.
[42] J. Liu, C. Wang, and S. Liu, "Utility of ChatGPT in clinical practice," *Journal of Medical Internet Research*, vol. 25, p. e48568, Jun. 2023, doi: 10.2196/48568.
[43] P. Giannos and O. Delardas, "Performance of ChatGPT on UK standardized admission tests: insights from the BMAT, TMUA, LNAT, and TSA examinations," *JMIR Medical Education*, vol. 9, p. e47737, Apr. 2023, doi: 10.2196/47737.
[44] L. De Angelis *et al.*, "ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health," *Frontiers in Public Health*, vol. 11, Apr. 2023, doi: 10.3389/fpubh.2023.1166120.
[45] S. Sarsa, P. Denny, A. Hellas, and J. Leinonen, "Automatic generation of programming exercises and code explanations using large language models," in *ICER 2022 - Proceedings of the 2022 ACM Conference on International Computing Education Research*, Aug. 2022, vol. 1, pp. 27–43, doi: 10.1145/3501385.3543957.
[46] H. Lee, "The rise of ChatGPT: exploring its potential in medical education," *Anatomical Sciences Education*, Mar. 2023, doi: 10.1002/ase.2270.
[47] X. Li, F. Tramèr, P. Liang, and T. Hashimoto, "Large language models can be strong differentially private learners," *ICLR 2022 - 10th International Conference on Learning Representations*, 2022.
[48] V. Slovikovskaya and G. Attardi, "Transfer learning from transformers to fake news challenge stance detection (FNC-1) task," *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings*, pp. 1211–1218, 2020.
[49] J. Welbl *et al.*, "Challenges in detoxifying language models," in *Findings of the Association for Computational Linguistics: EMNLP 2021*, 2021, pp. 2447–2469, doi: 10.18653/v1/2021.findings-emnlp.210.
[50] D. Luitse and W. Denkena, "The great transformer: examining the role of large language models in the political economy of AI," *Big Data and Society*, vol. 8, no. 2, p. 205395172110477, Jul. 2021, doi: 10.1177/20539517211047734.

[51] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable ai: a review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, pp. 1–45, Dec. 2021, doi: 10.3390/e23010018.

[52] R. Yang, T. F. Tan, W. Lu, A. J. Thirunavukarasu, D. S. W. Ting, and N. Liu, "Large language models in health care: Development, applications, and challenges," *Health Care Science*, vol. 2, no. 4, pp. 255–263, Aug. 2023, doi: 10.1002/hcs2.61.

[53] B. Miloski, "Opportunities for artificial intelligence in healthcare and in vitro fertilization," *Fertility and Sterility*, vol. 120, no. 1, pp. 3–7, Jul. 2023, doi: 10.1016/j.fertnstert.2023.05.006.

[54] J. J. Bird and A. Lotfi, "Generative transformer chatbots for mental health support: a study on depression and anxiety," in *ACM International Conference Proceeding Series*, Jul. 2023, pp. 475–479, doi: 10.1145/3594806.3596520.

[55] H. Huang *et al.*, "ChatGPT for shaping the future of dentistry: the potential of multi-modal large language model," *International Journal of Oral Science*, vol. 15, no. 1, p. 29, Jul. 2023, doi: 10.1038/s41368-023-00239-y.

[56] S. Thapa and S. Adhikari, "ChatGPT, bard, and large language models for biomedical research: opportunities and pitfalls," *Annals of Biomedical Engineering*, vol. 51, no. 12, pp. 2647–2651, Dec. 2023, doi: 10.1007/s10439-023-03284-0.

[57] F. Nugen *et al.*, "Application of natural language processing in total joint arthroplasty: opportunities and challenges," *Journal of Arthroplasty*, vol. 38, no. 10, pp. 1948–1953, Oct. 2023, doi: 10.1016/j.arth.2023.08.047.

[58] Y. Balla, S. Tirunagari, and D. Windridge, "Pediatrics in artificial intelligence era: a systematic review on challenges, opportunities, and explainability," *Indian Pediatrics*, vol. 60, no. 7, pp. 561–569, 2023, [Online]. Available: https://www.springer.com/journal/13312.

[59] V. Liyanage and D. Buscaldi, "Detecting artificially generated academic text: the importance of mimicking human utilization of large language models," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13913 LNCS, 2023, pp. 558–565.

[60] L. Yan *et al.*, "Practical and ethical challenges of large language models in education: a systematic scoping review," *British Journal of Educational Technology*, vol. 55, no. 1, pp. 90–112, Jan. 2024, doi: 10.1111/bjet.13370.

[61] U. Shakir, J. L. Hess, M. James, and A. Katz, "Pushing ethics assessment forward in engineering: NLP-assisted qualitative coding of student responses," 2023, doi: 10.18260/1-2--43997.

[62] M. Kreminski and C. Martens, "Unmet creativity support needs in computationally supported creative writing," in *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022)*, 2022, pp. 74–82, doi: 10.18653/v1/2022.in2writing-1.11.

[63] E. Z. Chang and O. J. Mengshoel, "Generating natural language dialogues using large language models with adapters," *CEUR Workshop Proceedings*, vol. 3431, 2023.

[64] E. Bauer *et al.*, "Using natural language processing to support peer-feedback in the age of artificial intelligence: a cross-disciplinary framework and a research agenda," *British Journal of Educational Technology*, vol. 54, no. 5, pp. 1222–1245, Sep. 2023, doi: 10.1111/bjet.13336.

[65] Y. Zhang, D. Towey, and M. Pike, "Automated metamorphic-relation generation with ChatGPT: an experience report," in *Proceedings - International Computer Software and Applications Conference*, Jun. 2023, vol. 2023-June, pp. 1780–1785, doi: 10.1109/COMPSAC57700.2023.00275.

[66] U. Iqbal, "From knowledge management to intelligence engineering - a practical approach to building ai inside the law-firm using open-source large language models," *CEUR Workshop Proceedings*, vol. 3423, pp. 49–57, 2023.

[67] J. Bae, C. Yu Hung, and L. van Lent, "Mobilizing text as data," *European Accounting Review*, vol. 32, no. 5, pp. 1085–1106, Oct. 2023, doi: 10.1080/09638180.2023.2218423.

[68] S. Shah *et al.*, "Robust NLP for finance (RobustFin)," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2023, pp. 5884–5885, doi: 10.1145/3580305.3599211.

[69] T. Ahmed, S. Ghosh, C. Bansal, T. Zimmermann, X. Zhang, and S. Rajmohan, "Recommending root-cause and mitigation steps for cloud incidents using large language models," in *Proceedings - International Conference on Software Engineering*, May 2023, pp. 1737–1749, doi: 10.1109/ICSE48619.2023.00149.

[70] T. Kimura, "Assessment of personal values for data-driven human resource management," *Data Science Journal*, vol. 22, Jul. 2023, doi: 10.5334/dsj-2023-022.

[71] L. P. Osco, E. L. de Lemos, W. N. Gonçalves, A. P. M. Ramos, and J. Marcato Junior, "The potential of visual ChatGPT for remote sensing," *Remote Sensing*, vol. 15, no. 13, p. 3232, Jun. 2023, doi: 10.3390/rs15133232.

[72] S. Bashir, M. Abbas, M. Saadatmand, E. P. Enoiu, M. Bohlin, and P. Lindberg, "Requirement or not, that is the question: a case from the railway industry," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13975 LNCS, 2023, pp. 105–121.

[73] A. Morales-Garzón, G. M. Sánchez-Pérez, J. C. Sierra, and M. J. Martin-Bautista, "The promise of query answering systems in sexuality studies: current state, challenges and limitations," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14113 LNAI, 2023, pp. 39–49.

[74] M. F. Wong, S. Guo, C. N. Hang, S. W. Ho, and C. W. Tan, "Natural language generation and understanding of big code for AI-assisted programming: a review," *Entropy*, vol. 25, no. 6, p. 888, Jun. 2023, doi: 10.3390/e25060888.

[75] A. Ahmad, M. Waseem, P. Liang, M. Fahmideh, M. S. Aktar, and T. Mikkonen, "Towards human-bot collaborative software architecting with ChatGPT," in *ACM International Conference Proceeding Series*, Jun. 2023, pp. 279–285, doi: 10.1145/3593434.3593468.

[76] G. Yang, "Augmented intelligence: a collaboration of humans and machines," *YouTube*, 2024. https://www.youtube.com/watch?v=CYbKuDdmaYs (accessed Feb. 15, 2024).

[77] D. M. Howcroft and D. Gkatzia, "Most NLG is low-resource: here's what we can do about it," in *GEM 2022 - 2nd Workshop on Natural Language Generation, Evaluation, and Metrics, Proceedings of the Workshop*, 2022, pp. 336–350, doi: 10.18653/v1/2022.gem-1.29.

[78] A. Seker, E. Bandel, D. Bareket, I. Brusilovsky, R. S. Greenfeld, and R. Tsarfaty, "AlephBERT: language model pre-training and evaluation from sub-word to sentence level," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2022, vol. 1, pp. 46–56, doi: 10.18653/v1/2022.acl-long.4.

[79] M. Godbole, P. Dandavate, and A. Kane, "Temporal word meaning disambiguation using timeLMs," in *EvoNLP 2022 - 1st Workshop on Ever Evolving NLP, Proceedings of the Workshop*, 2022, pp. 55–60, doi: 10.18653/v1/2022.evonlp-1.8.

[80] F. Barreto, L. Moharkar, M. Shirodkar, V. Sarode, S. Gonsalves, and A. Johns, "Generative artificial intelligence: opportunities and challenges of large language models," in *Lecture Notes in Networks and Systems*, vol. 699 LNNS, 2023, pp. 545–553.

[81] X. Fang, A. Kalinowski, H. Zhao, Z. You, Y. Zhang, and Y. An, "Prompt design and answer processing for knowledge base construction from pre-trained language models (KBC-LM)," *CEUR Workshop Proceedings*, vol. 3274, pp. 55–66, 2022.

[82] A. Lu, H. Zhang, Y. Zhang, X. Wang, and D. Yang, "Bounding the capabilities of large language models in open text generation with prompt constraints," in *Findings of the Association for Computational Linguistics: EACL 2023*, 2023, pp. 1982–2008, doi: 10.18653/v1/2023.findings-eacl.148.
[83] D. Cheng *et al.*, "Foundations and applications in large-scale AI models: pre-training, fine-tuning, and prompt-based learning," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Aug. 2023, pp. 5853–5854, doi: 10.1145/3580305.3599209.
[84] X. Hu, Y. Tian, K. Nagato, M. Nakao, and A. Liu, "Opportunities and challenges of ChatGPT for design knowledge management," *Procedia CIRP*, vol. 119, pp. 21–28, 2023, doi: 10.1016/j.procir.2023.05.001.
[85] E. N. Crothers, N. Japkowicz, and H. L. Viktor, "Machine-generated text: a comprehensive survey of threat models and detection methods," *IEEE Access*, vol. 11, pp. 70977–71002, 2023, doi: 10.1109/ACCESS.2023.3294090.
[86] S. M. Goodman *et al.*, "LaMPost: design and evaluation of an AI-assisted email writing prototype for adults with Dyslexia," in *ASSETS 2022 - Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*, Oct. 2022, pp. 1–18, doi: 10.1145/3517428.3544819.
[87] J. J. Chen and J. C. Lin, "Artificial intelligence as a double-edged sword: wielding the POWER principles to maximize its positive effects and minimize its negative effects," *Contemporary Issues in Early Childhood*, vol. 25, no. 1, pp. 146–153, Mar. 2024, doi: 10.1177/14639491231169813.

BIOGRAPHIES OF AUTHORS

Rachel Edita O. Roxas, PhD is Professor of the Institute of Computer Science, College of Arts and Sciences, University of the Philippines Los Baños. She has a Ph.D. in Computer Science from the Australian National University, and a second Ph.D. in Education from the De La Salle University. She led the development of Natural Language Processing in the country working on Philippine languages. She has been selected as the ASEAN member state national finalist representing the Philippines in the senior scientist category of the ASEAN Women in Science 2022 awards. She has also been named among the top 1,000 scientists of the Philippines in the AD Scientific 2024. She can be contacted at email: roroxas2@up.edu.ph.

Reginald Neil C. Recario is an Assistant Professor of the Institute of Computer Science, College of Arts and Sciences, University of the Philippines Los Banos. He obtained both his Bachelor of Science in Computer Science and Master of Information Technology at the University of the Philippines Los Banos. His research interests focus on computer vision, natural language processing, and artificial intelligence. He can be contacted at email: rcrecario@up.edu.ph.