

Quarterly Newsletter

CHAIR
iDANAE

INTELLIGENCE · DATA · ANALYSIS · STRATEGY

2023

Large Language Models:
a new era for Artificial Intelligence



POLITÉCNICA

UNIVERSIDAD
POLITÉCNICA
DE MADRID

MS^o *Management
Solutions*
Making things happen

Introduction

The rapid emergence of large language models (LLM) has captured the attention of professionals across various industries. Possessing exceptional capabilities in understanding, generating, and processing text, these models have proven to be valuable assets for numerous tasks associated to language, including machine translation, sentiment analysis, and automated content creation, among others.

As large language models continue to evolve, their applications become progressively sophisticated. These models streamline processes and automate tasks that previously required human intervention, particularly in fields such as customer service, journalism, or content creation. Therefore, this emerging technology unlocks numerous business opportunities, driving cost reduction and optimizing revenue generation.

However, this technology is not free of challenges and risks. While it consistently improves efficiency and lowers operational expenses, it also gives rise to concerns regarding potential job displacement, data protection, biases, environmental issues, or the possible misuse of confidential information, among many others.

The objective of this paper is to provide a comprehensive understanding of large language models by examining their technical underpinnings and recent advancements. In addition, there is a discussion on the challenges they may pose, both from a technical and an ethical standpoint. Finally, some last insights about the current state and prospects of large language models in the professional landscape are summarized.





Definition and industry updates

A large language model (LLM) is a machine learning model trained on vast amounts of text data to understand and generate human-like natural language. Using deep learning techniques, these models are trained to capture complex language patterns. After setting and training very large amounts of parameters, a LLM is enabled to perform tasks such as answering questions, summarizing text, paraphrasing, translation, or spell and grammar checking.

The origins of LLM can be traced back to the introduction of transformers, which emerged as a groundbreaking neural network architecture. A transformer is a type of neural network architecture particularly well suited to handle sequential data, like text. This significant development was presented in the seminal research paper titled "Attention Is All You Need" [1], authored by a team of Google researchers. Subsequently, the field of LLM has experienced exponential growth and advancement. The following are some of the state-of-the-art models in the field of LLM that exemplify the latest advancements and breakthroughs in research and technology:

- ▶ **GPT-3.5 by OpenAI [2].** Researchers developed and implemented the concept of generative pre-training [3]. Although there is an abundance of large unlabeled text corpora, labeled data for learning specific NLP tasks is scarce, making it challenging for discriminatively trained models to perform adequately. Significant gains on these tasks can be achieved by generatively pre-training a language model on a diverse corpus of unlabeled text, followed by discriminative fine-tuning on each specific task.
- ▶ **GPT-4 by OpenAI [4].** GPT-4 is an extensive, multimodal model that processes both image and text inputs to produce text outputs. Precise information regarding the model's architecture and training is not disclosed due to concerns about competition and security. GPT-4 outperforms its predecessors on standard benchmarks, demonstrating considerable advancements in comprehending user intentions and ensuring safety features. Additionally, the model attains human-equivalent performance on numerous tests, such as scoring within the top 10% on a simulated Uniform Bar Examination.
- ▶ **LLaMA by Meta [5].** LLaMA (Large Language Model Meta AI) is a collection of language models with 7B (billion) to 65B parameters. LLaMA 33B and 65B were trained on 1.4 trillion tokens (i.e., a unit of text the model processes, it can be a single character of a word), while the smallest model, LLaMA 7B, was trained on one trillion tokens. They exclusively used publicly available datasets, without depending on proprietary or restricted data. The Meta AI team asserts that smaller models trained on more tokens are easier to retrain and fine-tune for specific product applications. LLaMA-13B outperformed GPT-3 while being over 10 times smaller.
- ▶ **Bloom [6]** is a 176B parameter model developed collaboratively by over 1,000 AI researchers from various companies, including Hugging Face, Microsoft, and NVIDIA. As most LLM are created by resource-rich organizations and often remain inaccessible to the public, the authors aimed to democratize this powerful technology by releasing Bloom as an open-source model.
- ▶ **PaLM by Google [7].** Pathways Language Model (PaLM) is a 540B parameter, Transformer-based language model. PaLM was trained using Pathways, a new ML system for efficient training. The model demonstrates the benefits of scaling in few-shot learning, achieving state-of-the-art results on hundreds of language understanding and generation benchmarks. PaLM outperforms finetuned state-of-the-art models on multi-step reasoning tasks and exceeds average human performance on the BIG-bench benchmark. Recently, Google has developed an enhanced version of their model, PaLM 2.c
- ▶ **LaMDA by Google [8].** Language Models for Dialog Applications (LaMDA) is a series of Transformer-based models which have up to 137B parameters and pre-trained on 1.56T words of public data. LaMDA is capable to generate text from text input as well as images or structured data. Google's chatbot Bard is based on this model.
- ▶ **Claude by Anthropic [9].** This is a 52B parameter, pre-trained LLM that uses a reinforcement learning model, rather than humans, to generate the initial ranking of fine-tuned models. This LLM is the choice of AWS as a foundation model in its Bedrock service.

Technical review

The performance of an LLM can vary based on its architecture or type, with each designed to excel in certain tasks. For instance, encoder models (a type of neural network that transforms information from one representation to another one, with the purpose of compressing it) are more proficient in sentence classification, named entity recognition, and extractive question answering. Conversely, decoder models (a neural network executing the reverse of the encoder) demonstrate superior performance in text generation. Encoder-decoder models, on the other hand, are optimally designed for tasks such as summarization, translation, and generative question answering.

Transformers are at the core of state-of-the-art LLM. Transformers are based on 3 key components:

- ▶ **Self-attention mechanism:** It allows the model to weigh the importance of different parts of the input sequence for a specific output position. In other words, it helps the model to focus on relevant input tokens when generating output tokens. This is done using attention scores calculated from the input tokens' embeddings, which are then used to compute a weighted sum of the input representations. The self-attention mechanism can be applied multiple times in parallel, resulting in multi-head attention.
- ▶ **Positional encoding:** The Transformer architecture does not have an inherent sequential processing, like recurrent neural networks (RNN); instead, it incorporates the position information of tokens in the input sequence in a novel way. Positional encoding is added to the input embeddings to provide this information. Transformers use a smart positional encoding scheme, where each position/index is mapped to a vector. Hence, the output of the positional encoding layer is a matrix, where each row of the matrix represents an encoded object of the sequence summed with its positional information. This can be done using sinusoidal functions or learnable embeddings, allowing the model to understand the relative positions of tokens in the sequence.
- ▶ **Layer normalization:** Transformers consist of multiple layers of self-attention and feed-forward neural networks. Layer normalization is used to stabilize the training process by normalizing the activations of each layer before they are passed on to the next layer. This helps mitigate the vanishing or exploding gradient problem, which can occur during the training of deep neural networks.

The performance of LLM is typically measured using various benchmark tasks and datasets that evaluate the model's ability to understand and generate natural language. Performance on these tasks is measured using different evaluation metrics that quantify the models' accuracy, fluency, and relevance.

Some common evaluation metrics include:

- ▶ **Perplexity:** a measure of how well a language model predicts a given sequence of words.
- ▶ **BLEU (Bilingual Evaluation Understudy):** a metric commonly used to evaluate machine translation tasks. It measures the similarity between the model-generated translation and the human reference translation by comparing n-grams.
- ▶ **Precision, Recall and F1 score:** used to evaluate tasks like named entity recognition and question-answering. Low precision indicates a high number of false positives and low recall indicates a high number of false negatives. The F1 score conveys the balance between the precision and the recall.
- ▶ **ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [10]:** ROUGE is a set of metrics used for evaluating automatic text summarization. It measures the overlap between the generated summary and the human reference summary in terms of n-grams, word sequences, or other linguistic units.
- ▶ **Accuracy:** ratio of correct predictions to the total number of predictions made by the model. It is a straightforward metric used for classification tasks like sentiment analysis and natural language inference.
- ▶ **Character/Word/Sentence Error Rate (CER, WER, SER):** these metrics can also add to the analysis of the performance in terms of the error rate at character, word, or sentences (although these are metrics commonly used in automatic transcription tasks).

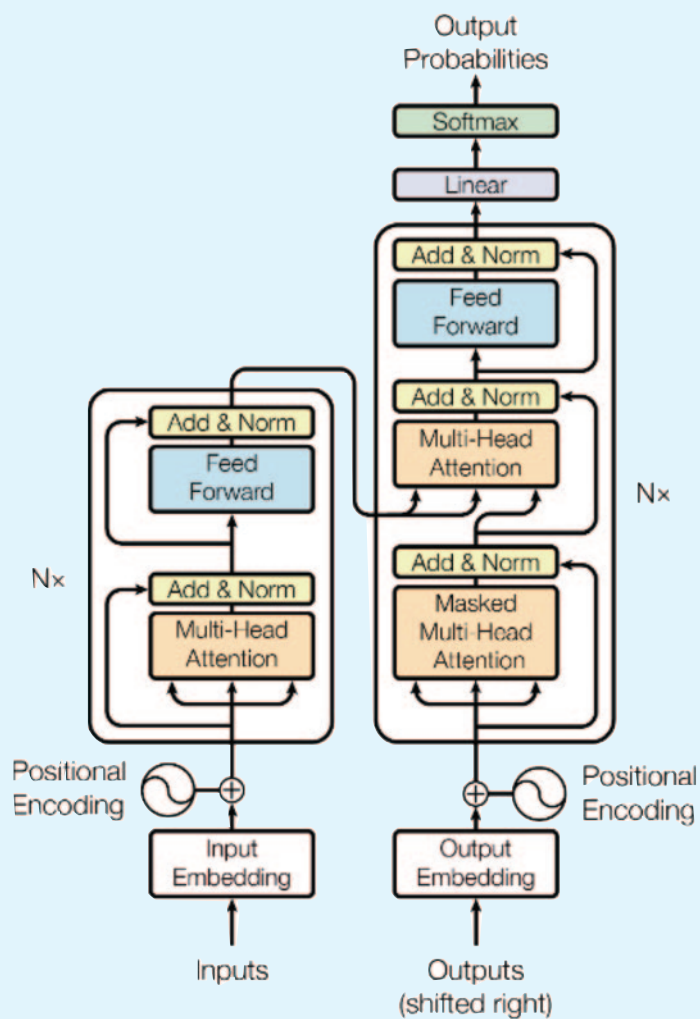
There are several benchmark datasets and leaderboards that facilitate the comparison of LLM based on their performance across various tasks:

- ▶ **GLUE [11] (General Language Understanding Evaluation):** GLUE is a benchmark of nine NLP tasks, including sentiment analysis, natural language inference, and paraphrase detection, aimed at evaluating the general understanding of language models.



- ▶ SuperGLUE [12]: This is an extension of the GLUE benchmark that introduces more challenging tasks and diverse datasets to assess the models' performance further.
- ▶ SQuAD [13] (Stanford Question Answering Dataset): SQuAD is a benchmark for evaluating reading comprehension and question-answering abilities of language models. It consists of a dataset containing questions and answers based on Wikipedia articles.
- ▶ LAMBADA (Language Modeling Broadened to Account for Discourse Aspects) [14]: LAMBADA is a language modeling task designed to evaluate the models' ability to predict the final word in a sentence, requiring the understanding of long-range dependencies and context.
- ▶ BIG-bench [15]: This dataset currently consists of 204 tasks, contributed by 444 authors across 132 institutions. Task topics are diverse, drawing problems from linguistics, math, common-sense reasoning, biology, physics, social bias, software development, and beyond. BIG-bench focuses on tasks that are believed to be beyond the capabilities of current language models.

Figure 1: Original transformer architecture [1]



Practical applications

LLM have demonstrated remarkable capabilities in processing and generating human-like text across various domains. These versatile AI models can perform a wide array of tasks, transforming the way we interact with and utilize textual data. The following list provides an overview of some of the key tasks that LLM can accomplish, highlighting their potential to transform numerous applications in both business and research settings:

- ▶ **Text generation:** Creating coherent and contextually relevant text based on a given prompt or input.
- ▶ **Sentiment analysis:** Identifying the sentiment or emotion (e.g., positive, negative, neutral) expressed in a piece of text.
- ▶ **Named entity recognition:** Identifying and classifying entities such as names, organizations, locations, and dates within a text.
- ▶ **Text summarization:** Generating a concise summary of a longer document or piece of text while preserving its essential meaning.
- ▶ **Question answering:** Providing accurate and relevant answers to user queries based on a given context or knowledge source.
- ▶ **Machine translation:** Converting text from one language to another while preserving its meaning and context.
- ▶ **Text classification:** Categorizing text into predefined classes or topics based on its content.
- ▶ **Keyword extraction:** Identifying and extracting significant keywords and phrases from a given text.
- ▶ **Paraphrasing:** Rewriting a piece of text in different words while retaining its original meaning.
- ▶ **Chatbot conversations:** Conducting natural and context-aware conversations with users.
- ▶ **Semantic search:** Improving search results by understanding the meaning and intent of a query, rather than relying solely on keyword matching.



- ▶ **Content filtering:** Identifying and filtering out inappropriate, offensive, or irrelevant content.
- ▶ **Natural language text generation from structured data:** Creation of written reports from sales or financial data.
- ▶ **Coding tasks:** generating code, translating between different languages, documenting and reviewing code.

As LLM continue to advance and grow more sophisticated, their practical applications are expanding across diverse industries, streamlining processes, and automating tasks that once required human intervention. The following list highlights a range of business use cases where LLM can be effectively employed, demonstrating their potential to enhance efficiency, save resources, and drive growth across various sectors:

1. Automating customer service

- ▶ **Chatbots:** Intelligent chatbots could understand user queries and provide accurate responses, improving customer service quality and efficiency.
- ▶ **Virtual assistants:** AI-powered virtual assistants can handle routine tasks such as scheduling appointments, answering FAQs, and providing general support, freeing up human agents to handle more complex tasks.
- ▶ **Helpdesk ticketing:** Analyzing incoming helpdesk tickets, categorize them based on urgency and subject matter, and route them to the appropriate team or agent for a quicker resolution.

2. Content creation

- ▶ **Marketing materials:** Generating persuasive and engaging copy for advertising campaigns, email newsletters, and promotional materials, helping businesses reach their target audience more effectively.
- ▶ **Social media content:** Creating compelling social media posts that resonate with a company's target audience, increasing brand awareness and engagement.
- ▶ **Product descriptions:** Producing accurate and persuasive product descriptions that showcase the benefits and features of products, improving conversion rates and sales.

3. Market research

- ▶ **Sentiment analysis:** Analyzing customer reviews, social media posts, and other online content to gauge public sentiment towards a brand or product, helping businesses

make data-driven decisions and improve customer satisfaction.

- ▶ **Social media monitoring:** AI-driven tools can track brand mentions, industry trends, and relevant conversations on social media platforms, enabling businesses to identify opportunities and threats in real-time.
- ▶ **Competitor analysis:** Extracting valuable insights from competitors' websites, marketing materials, and online content, helping businesses understand their competitive landscape and develop effective strategies to stand out.

4. Personalization and recommendation systems

- ▶ **Personalized content:** By analyzing user behavior, preferences, and browsing history, LLM can generate personalized content that caters to individual users, increasing engagement and user satisfaction.
- ▶ **Targeted advertising:** Creating tailored marketing messages for specific audience segments, improving ad relevance and return on investment.

5. Legal and compliance

- ▶ **Contract analysis:** Reviewing legal contracts, identifying key terms, clauses, and potential risks, saving time, and reducing the likelihood of costly oversights.
- ▶ **Regulatory compliance monitoring:** LLM can help organizations stay compliant by monitoring regulatory changes and providing insights on how new rules or updates may impact their business operations, reducing the risk of non-compliance penalties.

6. Healthcare and medical research

- ▶ **Medical record analysis:** Analyzing electronic health records, extracting relevant information, and identifying patterns that could help healthcare professionals make better-informed decisions and improve patient outcomes.

7. Human resources

- ▶ **Resume screening:** Scanning resumes, identifying the most relevant candidates based on skills, experience, and other factors, streamlining the recruitment process, and reducing hiring bias.
- ▶ **Job matching:** Matching job seekers with suitable job opportunities by analyzing their resumes and job preferences, and comparing them to available positions, improving the overall job search experience.

8. Finance

- ▶ Financial news analysis: Monitoring and analyzing financial news, identifying market-moving events, trends, and sentiments that may impact investment decisions and portfolio performance.
- ▶ Sentiment-based trading strategies: LLM can be used to develop trading strategies based on the sentiment analysis of financial news, social media, and other sources, potentially improving investment returns.

9. Education and training

- ▶ Learning assistant: Explaining complex concepts in simple terms, provide answers to frequently asked questions, and offer personalized learning resources on online education platforms.
- ▶ Creation of educational content: Developing study materials, practice exams, and other learning resources.

10. Business processes automation

- ▶ Report automation: Generating written reports from structured data, such as sales, performance metrics, or financial data.
- ▶ Email workflow: Categorize and respond to routine emails, allowing workers to focus on more important tasks.

An example of use: Improving data extraction from unstructured documents

In this example, a potential use of LLM in a real-world business application is showcased, exemplifying how these advanced AI models can be harnessed to solve complex problems and improve existing processes.

In relation to the recruiting process, large companies typically receive hundreds, or even thousands, of resumes daily, that must be processed for the decision making with regards to hiring. Traditionally, these documents undergo an initial review by a human who filters candidates based on their education or experience. The selected resumes then receive a more in-depth evaluation as part of the ongoing selection process.

To perform this task in a more efficient way, companies ask candidates to input their details into online forms or load the CV in a specific pre-defined format; however, this approach might deter applicants from performing the work, potentially causing organizations to overlook highly skilled and talented individuals. To solve this point, one alternative is to request just a CV, and then use models to extract the entities and the information on the different criteria.

To do this task, the following system would follow these steps:

- ▶ Convert the initial resume file (PDF, word, image) into a unified image file.
- ▶ Run and inspect the image through an optical character recognition (OCR) AI system. This type of AI technology transforms images into text.



- Use the extracted text from the resume by the OCR to infer relevant information about the candidate.

For obtaining the information in text form after applying the OCR, some fields need to be identified in the extraction phase, using techniques such as identity recognition. Specific information, such as the candidate's name (typically located at the top of the resume), phone number, or email address, can be readily extracted using regular expressions (a regular expression is a search pattern used for matching and manipulating text, allowing for tasks such as pattern matching, validation, text manipulation, and parsing).

However, in fields such as education, experience, and skills, the previous strategy cannot be applied since text data is usually unstructured and does not follow any specific pattern such as an email. Therefore, regular expression cannot be used. Since LLM enable the automatic extraction of information from text, LLM could be used to resolve this issue, either summarizing text segments or extracting information directly in the desired format. For instance, if in a CV the information on the education is given by the text "I studied at the University of Barcelona,

where I obtained a degree in mathematics", after applying the LLM the structured data output could be: "University of Barcelona" and "Bachelor's degree in Mathematics".

It could be presumed that searching the text for exact matches would be adequate; however, the roster of available universities and degrees is constantly evolving. Moreover, this approach would only be effective for resumes and matches in a single language. Cycling through the list of all available universities, across every language, would demand substantial computational power, leading to an unwieldy user experience. Therefore, integrating an LLM for this process would increase the efficiency, and improve the experience of the candidate in the selection process.



Limitations and future challenges

LLM, while powerful and capable of generating impressive results, have some limitations. Some of the main limitations include:

Issues related to the generated output:

- ▶ **Bias:** LLM are trained on vast amounts of text data from the internet, which may include biased, offensive, or controversial content. Consequently, these models may inadvertently learn and perpetuate these biases in their generated text.
- ▶ **Sensitivity to input phrasing:** The performance of LLM can sometimes be sensitive to the way questions or prompts are phrased. This means that slight rephrasing might yield different, and potentially more accurate, results.
- ▶ **Hallucinations:** a model is said to "hallucinate" when it generates information or details that were not present in the input data or in the provided context.
- ▶ **Difficulties with information verification:** LLM can generate responses based on language patterns rather than verified facts, which can lead to the propagation of incorrect or false information.
- ▶ **Requires supervision:** LLM, despite being autonomous to a certain extent, still require considerable human supervision and adjustment to ensure accuracy and avoid undesirable outcomes.

Issues related to the architecture:

- ▶ **Lack of understanding:** Language models can generate text that appears coherent and contextually appropriate, but they may not possess a deep understanding of the content. As a result, the generated text may sometimes be factually incorrect or nonsensical.
- ▶ **Inability to handle multi-modal data:** Although recent advances have started to bridge the gap, LLM are primarily designed for text data and may struggle to handle multi-modal data, such as images or audio, without additional architecture modifications.
- ▶ **Absence of long-term memory:** LLM may not maintain consistent tracking of past interactions, which can limit their ability to sustain coherent and contextual long-term conversations.

Computational related issues:

- 4 **Model size and computational requirements:** State-of-the-art LLM have billions of parameters, resulting in substantial computational requirements for both training and inference. This makes them resource-intensive and less accessible to users with limited computational power.
- ▶ **Environmental impact:** The energy consumption associated with training and running LLM contributes to their carbon footprint, raising concerns about their environmental impact.



Ethical issues:

- **Safety and misuse concerns:** The capabilities of LLM can be misused to generate harmful content, such as disinformation or offensive material, raising concerns about their ethical use and potential regulation.

Another concerning issue regarding LLM is their rapid advancement and unpredictable impact on the global job market. These models open the possibility of automating many tasks that people currently perform within a very short time span, making the reallocation of human resources a significant challenge.

OpenAI recently published “GPTs are GPTs: An Early Look at the Labour Market Impact Potential of Large Language Models” [16], a working paper where the authors investigate the potential implications of LLM on the U.S. labour market. They assessed occupations based on their alignment with LLM capabilities, integrating both human expertise and GPT-4 classifications. They found out that around 80% of the U.S. workforce could have at least 10% of their work tasks affected by the introduction of LLM, while approximately 19% of workers may see at least 50% of their tasks impacted.

The projected effects span all wage levels, with higher-income jobs potentially facing increased exposure to LLM capabilities and LLM-powered software. Notably, these impacts are not limited to industries with higher recent productivity growth. Researchers found that 15% of all worker tasks in the US could be completed significantly faster while maintaining the same level of quality. When incorporating software and tools built on top of LLM, this proportion rises to between 47 and 56% of all tasks. This finding suggests that LLM-powered software will play a significant role in amplifying the economic impacts of the underlying models.

Lastly, there is a growing concern regarding data protection. Italy has become the first Western country to ban ChatGPT [17]: The Italian Data Protection Authority has instructed OpenAI to temporarily halt the processing of Italian users' data due to an ongoing investigation into a potential violation of Europe's stringent privacy regulations. As of April 30th, 2023, ChatGPT became accessible again in Italy.

Companies that own LLM must comply with the General Data Protection Regulation (GDPR) if they wish to offer their services in European countries. One of the primary challenges these companies face is complying with Article 17, which pertains to the right to be forgotten. Under GDPR, individuals have the right to request the removal of their personal information from an organization's records.

AI systems, such as neural networks, do not forget in the same way humans do. Instead, the network adjusts its weights to better accommodate new data, resulting in different outcomes for identical inputs. This process does not constitute forgetting in the traditional sense; rather, the network prioritizes the latest data it acquires. All the information remains within the system when retraining the weights of the network.

All these limitations and challenges raise the need for companies to define and integrate a proper framework and policies for their use of LLM in their daily work, including the addition of these elements to the governance, the creation of specific policies, establishing internally accepted uses of LLM, training the employees, and ensuring the correct use of the tools.

Conclusions

Large Language Models (LLM) are deep learning-based models capable of processing vast amounts of text. They can perform various tasks, such as text summarization, sentiment analysis, and machine translation. The rapid development and popularization of these models has opened the door to numerous business applications, particularly in automating tasks that are currently performed by humans.

Generating content for social media, analysing market trends, screening resumes for human resources, and automating customer service through chatbots represent only a fraction of the potential applications of LLM. The scope of these applications is vast, spanning numerous business sectors where LLM can provide significant value.

The numerous benefits and applications of these models are also accompanied by notable limitations, such as model bias and environmental impacts linked to their use. In addition, the integration of these models into our society presents significant challenges. Foremost among these are the unpredictable effects on the job market, the associated concerns surrounding user data protection, or the environmental impact these models have.

All these elements makes that the incorporation of LLM in business processes must be carried out by ensuring its effective, ethical and secure use. This implies identifying the uses that the organisation is willing to undertake through this technology, guaranteeing the confidentiality of the company's information (avoiding information leaks to third-party servers and anonymising the information), and ensuring that the outputs of these models are checked with reliable sources before integrating them into decision-making or management processes.

Note: This publication was produced with the help of several artificial intelligence (AI) tools. These tools were used for various tasks, such as searching for information, collecting and organizing data, and generating summaries. In any case, the final material in this publication was written by a person and not by an AI.

Bibliography



- [1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, "Attention is All You Need", in Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017.
- [2] L. Floridi and M. Chiriatti, "GPT-3: Its Nature, Scope, Limits, and Consequences", *Minds and Machines*, vol. 30, no. 4, pp. 681-694, 2020.
- [3] A. Radford, K. Narasimhan, T. Salimans and I. Sutskever, "Improving Language Understanding by Generative Pre-Training", 2019.
- [4] OpenAI, "GPT-4 Technical Report," 2023.
- [5] Meta AI, "LLaMA: Open and Efficient Foundation Language Models", 2023.
- [6] BigScience Workshop, "BLOOM: A 176B-Parameter Open-Access Multilingual Language Model", 2022.
- [7] Google Research, "PaLM: Scaling Language Modeling with Pathways", 2022.
- [8] Google, "LaMDA: Language Models for Dialog Applications", 2022.
- [9] Anthropic, "Constitutional AI: Harmlessness from AI Feedback", 2022.
- [10] C.-Y. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries", in *Text Summarization Branches Out*, Barcelona, Association for Computational Linguistics, 2004, pp. 74-81.
- [11] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy and S. Bowman, "GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding", in *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, 2018.
- [12] A. Wang, Y. Pruksachatkun, N. Nangia, A. Singh, J. Michael, F. Hill, O. Levy and S. R. Bowman, "SuperGLUE: A Stickier Benchmark for General-Purpose Language Understanding Systems", in *NIPS'19: Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 2019.
- [13] [Online]. Available: <https://rajpurkar.github.io/SQuAD-explorer/>.
- [14] D. Paperno, G. Kruszewski, A. Lazaridou, N. Q. Pham, R. Bernardi, S. Pezzelle, M. Baroni, G. Boleda and R. Fernandez, "The LAMBADA dataset: Word prediction requiring a broad discourse context", in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2016.
- [15] Google, [Online]. Available: <https://github.com/google/BIG-bench>.
- [16] T. Eloundou, S. Manning, P. Mishkin and D. Rock, "GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models", 2023.
- [17] CNBC, [Online]. Available: <https://www.cnbc.com/2023/04/04/italy-has-banned-chatgpt-heres-what-other-countries-are-doing.html>. [Accessed 25 April 2023].

Authors

Ernestina Menasalvas (UPM)

Víctor Robles (UPM)

Manuel Ángel Guzmán (Management Solutions)

Sergio Ruiz Bonilla (Management Solutions)



POLITÉCNICA

UNIVERSIDAD
POLITÉCNICA
DE MADRID

The Universidad Politécnica de Madrid is a multi-sector and multi-disciplinary Public Law Entity, which carries out teaching, research and scientific and technological development activities.

www.upm.es



Management Solutions is an international consulting firm, focused on business, finance, risk, organization, technology and process consulting, operating in more than 50 countries and with a team of over 3,300 professionals working for more than 1,500 clients worldwide.

www.managementsolutions.com

For more information visit

blogs.upm.es/catedra-idanae/