



The Artificial Intelligence Newsletter

August 2023



Introduction

Staying ahead of the curve

At PwC Albania and Kosovo, we are always on the lookout for new technologies, trending topics, inside stories on the future of technology and innovations globally - aiming to share them with all tech enthusiasts.

The first edition of the AI Newsletter intended to summarize latest trends in artificial intelligence, including transformers, generative models, large language models (LLMs), natural language processing (NLP), and diffusion models.

The second edition covered the topics of predictive analytics, automated decision making, big data, constructing the world of AI models using descriptive data, and related modern age ethical concerns of AI that must be taken into account as AI technology advances and becomes more prevalent in society.

The third edition focused on the potential positive and negative impacts of AI, potential political and regulatory solutions as well as its impact on the future of work and wealth.

In our fourth edition we introduce some of the most impactful industry players, as well as language model benchmarking best performers.

Our aim is to provide our readers with insights and updates on the latest developments in the field of AI and to continue to share our passion for technology with all tech enthusiasts.

*Through these newsletters PwC Albania and Kosovo is not endorsing any of the technologies or solutions mentioned, but rather summarizing and bringing to readers trending technologies which are defined by external experts as game-changing technologies.

Impactful players in the industry



Companies to keep an eye on in the race for AI

1. **NVIDIA** is a technology company that designs GPU technology and that has made significant contributions to the field of AI, including the development of hardware and software solutions for deep learning and other AI applications. NVIDIA's GPUs and its CUDA platform, a parallel computing platform for AI and high-performance computing, are used for running complex machine learning models. NVIDIA delivers the world's fastest AI training performance among commercially available products, according to MLPerf Benchmarks. Therefore, as stated in their blog, they were chosen by companies and institutions such as Amazon, Baidu, Facebook, MIT, and Stanford University for their AI initiatives. NVIDIA's hardware and software solutions enable organizations to build and deploy AI applications more efficiently.
2. **Microsoft Corporation** is a technology firm with an expertise in AI. As stated in Microsoft news, over the years, the company has developed a range of AI products and services, including Cortana, Microsoft Dynamics 365, Microsoft Teams, and has been involved in AI research and development. Microsoft is dedicated towards responsible AI and has established an AI ethics board to guide its AI initiatives. With its commitment to responsible AI and broad range of AI products and services, Microsoft plays a significant role in the future of AI. As it appears in Forbes News, Microsoft confirms its \$10 billion investment in OpenAI, enabling the company to leverage OpenAI's innovations and integrate them into its consumer products, while providing access to its technical infrastructure to host and run large AI models.
3. **Baidu Inc.** is a Chinese technology company specializing in AI research and development. The company offers a range of AI products and services, including Baidu Maps, Baidu Translate, Baidu Duer, and more. According to the recent IDC report "China AI Public Cloud Service Market Share, 2022", shared by PR Newswire - Baidu News, the company ranks first in China among others. The report highlights the continuous high-speed expansion of the AI public cloud service market in China, driven by Baidu's successful efforts in implementing intelligent solutions across diverse industries including manufacturing, energy, finance, and public service. Baidu has an infrastructure that enables it to host and run complex machine learning models at scale, demonstrating its commitment to advancing AI.

Impactful players in the industry



Companies to keep an eye on in the race for AI

4. **OpenAI** - As indicated by [OpenAI](#), the company is a research organization in the field of artificial intelligence, with a focus on creating responsible AI systems for the benefit of humanity. The company has made contributions to the development of advanced AI technologies such as deep learning, reinforcement learning, and generative models, and its research findings have been adopted by other organizations in the AI industry. OpenAI is also involved in exploring the ethical and societal implications of AI, making it a resource for those interested in the responsible development and deployment of AI.
5. **Google** - Google is a global technology company that has been involved in AI research and development for many years. As mentioned on the [Medium article](#), the company has developed a number of AI products, such as Google Assistant, Google Photos, and Google Translate, and has also made investments in AI-related startups. Google is also a contributor to the open-source AI community, with its TensorFlow library being a widely used machine learning framework. Google also owns DeepMind - a research organization in the field of AI, with a focus on creating general AI systems that can perform a wide range of tasks.
6. **Facebook** has also contributed to the field of AI through its research and development initiatives. One of the tools developed by Facebook's AI Research lab is PyTorch, an open-source machine learning library. According to [MetaAI](#), PyTorch is based on the Torch library and provides a flexible and dynamic computational graph, which allows for easy and intuitive model building. The library has become a tool for the AI and machine learning community due to its ease of use, features, and support from the community. With the recent release of PyTorch Meta, a new framework for meta-learning, Facebook continues to advance the field of AI and provide the community with resources.

Language Model Benchmarking on LAMBADA



Which is the best performer?

The LAMBADA (Language Modeling Broadened to Account for Discourse Aspects) benchmark is an open-ended cloze task which consists of about 10,000 passages from BooksCorpus where a missing target word is predicted in the last sentence of each passage. The missing word is constrained to always be the last word of the last sentence and there are no candidate words to choose from. Examples were filtered by humans to ensure they were possible to guess given the context, i.e., the sentences in the passage leading up to the last sentence. Examples were further filtered to ensure that missing words could not be guessed without the context, ensuring that models attempting the dataset would need to reason over the entire paragraph to answer questions.

Rank	Model	Accuracy	Perplexity	Paper	Year	Type
1	PaLM-540B (Few-Shot)	89.7	1.92	PaLM: Scaling Language Modeling with Pathways	2022	few-shot
2	Megatron-Turing NLG 530B (Few-Shot)	87.2		Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, A Large-Scale Generative Language Model	2022	few-shot
3	GPT-3 175B (Few-Shot)	86.4		Language Models are Few-Shot Learners	2020	few-shot
4	Cohere Large	82.33			2021	
5	PaLM-540B (One-Shot)	81.8		PaLM: Scaling Language Modeling with Pathways	2022	one-shot
6	GLaM 62B/64E (One-Shot)	80.9		GLaM: Efficient Scaling of Language Models with Mixture-of-Experts	2021	one-shot
7	GLM-130B (bidirectional attention)	80.2		GLM-130B: An Open Bilingual Pre-trained Model	2022	
8	SparseGPT (175B, 2:4 Sparsity)	79.47		SparseGPT: Massive Language Models Can Be Accurately Pruned in One-Shot	2023	
9	SparseGPT (175B, 4:8 Sparsity)	78.77		SparseGPT: Massive Language Models Can Be Accurately Pruned in One-Shot	2023	
10	PaLM-540B (Zero-Shot)	77.9		PaLM: Scaling Language Modeling with Pathways	2022	zero-shot
11	Chinchilla (Zero-Shot)	77.7		Training Compute-Optimal Large Language Models	2022	zero-shot

*Source: [link](#)

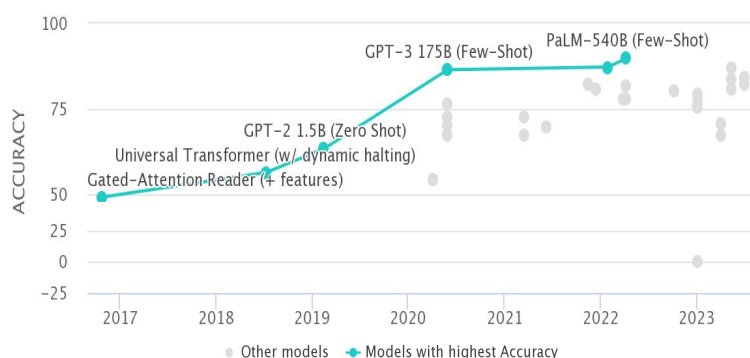
Language Model Benchmarking on LAMBADA



Which is the best performer?

Every year, researchers build new datasets with the goal of benchmarking new algorithms on something that existing algorithms are not able to solve. While several benchmarks are still unsolved after many years, other benchmarks face surprising progress over time.

The image on the right shows the accuracy reached by NLP models over the years on the LAMBADA dataset.



Rank	Model	Accuracy	Perplexity	Paper	Year	Type
12	SparseGPT (175B, 50% Sparsity)	76.51		SparseGPT: Massive Language Models Can Be Accurately Pruned in One-Shot	2023	
13	GPT-3 175B (Zero-Shot)	76.2	3	Language Models are Few-Shot Learners	2020	zero-shot
14	OPT-175B	75.59		SparseGPT: Massive Language Models Can Be Accurately Pruned in One-Shot	2023	
15	GPT-3 13B (Zero-Shot)	72.5	3.56	Language Models are Few-Shot Learners	2020	zero-shot
16	GLM-XXLarge (bidirectional)	72.35		GLM: General Language Model Pretraining with Autoregressive Blank Infilling	2021	zero-shot
17	GPT-3 6.7B (Zero-Shot)	70.3	4	Language Models are Few-Shot Learners	2020	zero-shot
18	GPT-J-6B	69.7	3.99		2021	
19	GLM-XXLarge (unidirectional)	67.18		GLM: General Language Model Pretraining with Autoregressive Blank Infilling	2021	zero-shot
20	GPT-3 2.7B (Zero-Shot)	67.1	4.6	Language Models are Few-Shot Learners	2020	zero-shot
21	GPT-2 1.5B (Zero Shot)	63.24	8.63	Language Models are Unsupervised Multitask Learners	2019	zero-shot
22	Universal Transformer (w/ dynamic halting)	56.25		Universal Transformers	2018	
23	Residual Shuffle-Exchange network	54.34		Residual Shuffle-Exchange Networks for Fast Processing of Long Sequences	2020	

*Source: [link](#)



Thank you for following our series of newsletters dedicated to AI. Stay tuned for more upcoming news and articles!

For more information, contact us:

PricewaterhouseCoopers Albania
Str. "Ibrahim Rugova", Sky Tower, 9/1

Telephone: +355 4 22 90 700

Email: al_pwc_albania@pwc.com

Web: www.pwc.com/al

