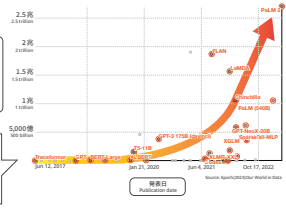
生成AIのインパクトはどこから来たのか?
Where Did the Impacts of Generative AI Come from?

生成AIは社会にさまざまな影響をおよぼしています。この影響の背後に連なる技術や要因のつながりをキーワードでふりかえります。

Generative AI is having various impacts on society. We will review the connections of technologies and factors behind this influence through keywords.

生成AIの学習につかわれるデータセットのサイズはここ数年で大幅に増加した。

The size of the datasets used for training generative AI has significantly increased.



AIをつくる人もおどろいた!
AI研究開発の加速

Even the researchers of AI were surprised! Acceleration of AI research and development.

研究開発への投資過熱
"Investment frenzy" in AI research and development

投資へのリターンが、他の技術と比べてもより確実に見通せるため投資が集まった。

Investment poured in due to the more predictable returns compared to other technologies.

スケーリング則の発見
Discovery of Scaling law

学習データを飛躍的にふやしていった結果、ある条件を満たせばスケーリング則に従って指数関数的に性能が向上することが発見された。[1]

As a result of greatly increasing the training data, it was discovered that under certain conditions, performance can improve exponentially following Scaling law.

自己教師あり学習(穴埋め問題)の活用
Utilization of Self-supervised learning (fill-in-the-blanks problems)

人間が教師データをつくることがボトルネックだったが、ウェブ上の莫大な文章をそのまま利用し、AI自身がその一部を穴あきにして、元の文章を教師データとすることで学習データをどんどんふやすことができた。

Creating human-labeled training data was a bottleneck. By using vast amounts of text from the web, AI could leave parts of blank to create gapped sentences, and then use the original text as teacher data. This allowed the training data to be increased dramatically.

炎上リスクの回避
Mitigation of controversy risks

生成AIの出力を多くの人間にとって好ましいものに制御できるようになり、開発企業は、不適切な出力によって非難を受け投資やユーザーが離れてしまうリスクを低減できた。

Generating AI's outputs to be controllable was favored to many people, and development companies were able to reduce the risk of criticism, investment withdrawal, and user attrition caused by inappropriate outputs.

RLHF(人間からのフィードバックによる強化学習)によるしつけ
Breeding by Reinforcement Learning from Human Feedback (RLHF)

学習データにふくまれるバイアスが再現されてしまうと、攻撃的であったり偏見をふくんだ文章を出力してしまう。チャットGPTでは、ラベラーとよばれる専門家が教師となり多くの人にとって望ましい回答を学習させた。[2]

If biases present in the training data are replicated, the generated outputs can become offensive or biased. In the case of ChatGPT, experts called "labelers" were used as teachers to train the model to provide responses that are desirable to a wide range of people.

利用者が爆発的に増えた!
臨機応変で自然な受け答え

Generated an explosive increase in users! Flexible and natural responses in real-time

コンテキスト内学習(文脈における学習)の創発
Emergence of In-context learning

話題や状況といった文脈に則した返答の方法を、生成AIのモデルをチューニングすることなくユーザーとの対話から学習できるようになった。これは研究者も意図していなかった。[3]

The ability to generate contextually appropriate responses based on topics and situations, without the need to fine-tune generative AI, has been achieved through learning from interactions with users. This was an outcome that even the researchers had not initially anticipated.

自己注意機構の副産物?
A byproduct of Self-attention mechanism?

コンテキスト内学習をする多くの生成AIは、Transformerという深層学習モデルをもとにしている。Transformerのもつ自己注意機構は、ユーザーの入力や生成しているテキストの文脈に合わせて、モデル自体をチューニングしたときのような挙動の調整を実現していると考えられている。[4]

Many generative AIs that perform in-context learning are based on a deep learning model called Transformer. The self-attention mechanism inherent in Transformer allows the model to adjust its behavior. This is similar to "tuning" itself when taking into account user inputs and the context of the generated text.

注釈

[参考文献]
岡野原大輔 (2023)『大規模言語モデルは新しい知能か』岩波書店
山田育矢ほか (2023)『大規模言語モデル入門』技術評論社

1 Kaplan, J. et al. (2020). Scaling Laws for Neural Language Models. *arxiv preprint*. arXiv:2001.08361

2 Ouyang, L. et al. (2022). Training language models to follow instructions with human feedback. *arxiv preprint*. arXiv:2203.02155

3 Brown, T. B. et al. (2020). Language Models are Few-Shot Learners. *arxiv preprint*. arXiv:2005.14165

4 Oswald, J. V. et al. (2023). Transformers learn in-context by gradient descent. *arxiv preprint*. arXiv:2212.07677