Scientific writing

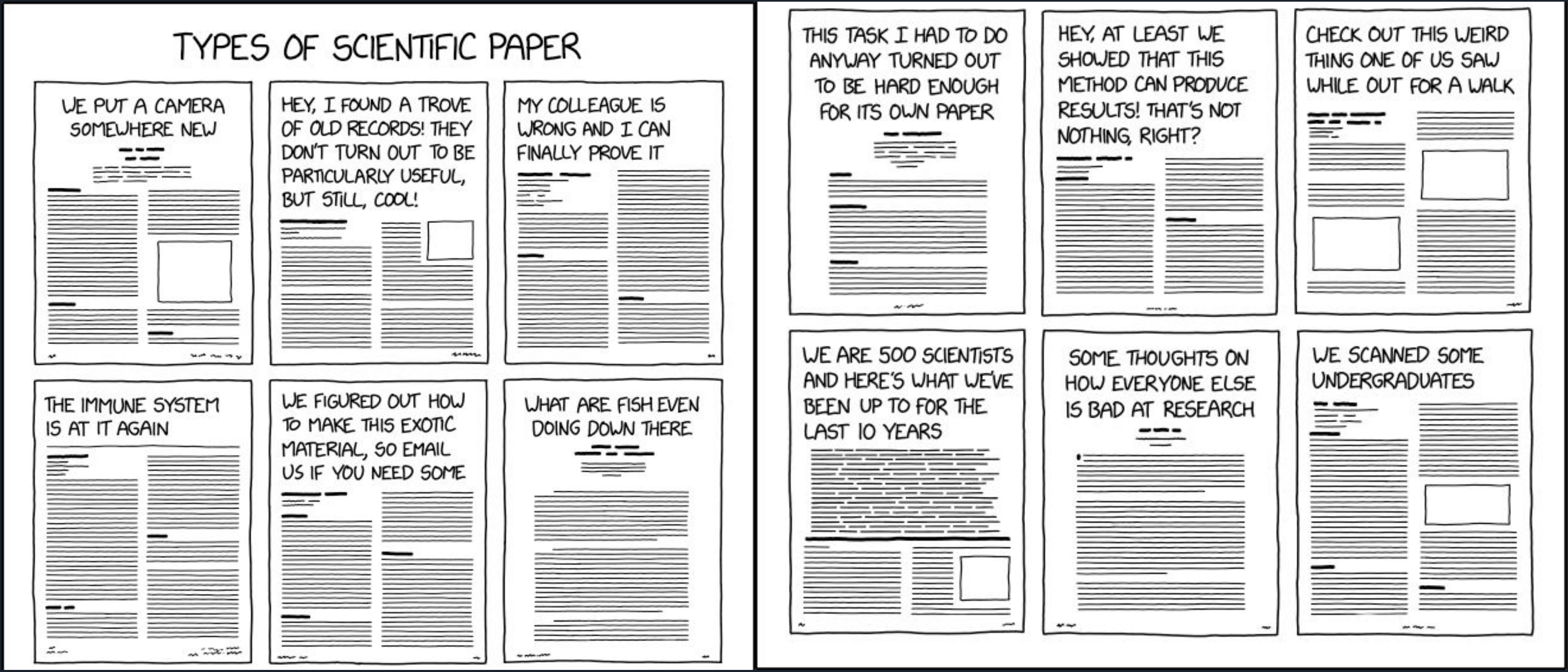
Reporting of original research in journals through scientific papers that follow a standard format.

The key characteristic of scientific writing is clarity.

Successful scientific experimentation is the result of a attacking a clearly stated problem and producing clearly stated conclusions.

Scientific paper

A scientific paper is a written, peer reviewed, and published report describing original research results.


TYPES OF SCIENTIFIC PAPER
WE PUT A CAMERA SOMEWHERE NEW
HEY, I FOUND A TROVE OF OLD RECORDS! THEY DON'T TURN OUT TO BE PARTICULARLY USEFUL, BUT STILL, COOL!
MY COLLEAGUE IS WRONG AND I CAN FINALLY PROVE IT
THE IMMUNE SYSTEM IS AT IT AGAIN
WE FIGURED OUT HOW TO MAKE THIS EXOTIC MATERIAL, SO EMAIL US IF YOU NEED SOME
WHAT ARE FISH EVEN DOING DOWN THERE
THIS TASK I HAD TO DO ANYWAY TURNED OUT TO BE HARD ENOUGH FOR ITS OWN PAPER
HEY, AT LEAST WE SHOWED THAT THIS METHOD CAN PRODUCE RESULTS! THAT'S NOT NOTHING, RIGHT?
CHECK OUT THIS WEIRD THING ONE OF US SAW WHILE OUT FOR A WALK
WE ARE 500 SCIENTISTS AND HERE'S WHAT WE'VE BEEN UP TO FOR THE LAST 10 YEARS
SOME THOUGHTS ON HOW EVERYONE ELSE IS BAD AT RESEARCH
WE SCANNED SOME UNDERGRADUATES

Importance of Papers - AI Trends

arXiv:2402.17177v2 [cs.CV] 28 Feb 2024

Sora: A Review on Background, Technology, Limitations, and Opportunities of Large Vision Models

Yixin Liu[1*] Kai Zhang[1*] Yuan Li[1*] Zhiling Yan[1*] Chujie Gao[1*]
Ruoxi Chen[1*] Zhengqing Yuan[1*] Yue Huang[1*] Hanchi Sun[1*]
Jianfeng Gao[2] Lifang He[1] Lichao Sun[1†]

[1]Lehigh University [2]Microsoft Research

Abstract

Warning: This is not an official technical report from OpenAI.
Sora is a text-to-video generative AI model, released by OpenAI in February 2024. The model is trained to generate videos of realistic or imaginative scenes from text instructions and show potential in simulating the physical world. Based on public technical reports and reverse engineering, this paper presents a comprehensive review of the model's background, related technologies, applications, remaining challenges, and future directions of text-to-video AI models. We first trace Sora's development and investigate the underlying technologies used to build this "world simulator". Then, we describe in detail the applications and potential impact of Sora in multiple industries ranging from film-making and education to marketing. We discuss the main challenges and limitations that need to be addressed to widely deploy Sora, such as ensuring safe and unbiased video generation. Lastly, we discuss the future development of Sora and video generation models in general, and how advancements in the field could enable new ways of human-AI interaction, boosting productivity and creativity of video generation.

Figure 1: Sora: A Breakthrough in AI-Powered Vision Generation

DeepMind

Training Compute-Optimal Large Language Models

Jordan Hoffmann*, Sebastian Borgeaud*, Arthur Mensch*, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, Tom Hennigan, Eric Noland, Katie Millican, George van den Driessche, Bogdan Damoc, Aurelia Guy, Simon Osindero, Karen Simonyan, Erich Elsen, Jack W. Rae, Oriol Vinyals and Laurent Sifre*

*Equal contributions

We investigate the optimal model size and number of tokens for training a transformer language model under a given compute budget. We find that current large language models are significantly under-trained, a consequence of the recent focus on scaling language models whilst keeping the amount of training data constant. By training over 400 language models ranging from 70 million to over 16 billion parameters on 5 to 500 billion tokens, we find that for compute-optimal training, the model size and the number of training tokens should be scaled equally: for every doubling of model size the number of training tokens should also be doubled. We test this hypothesis by training a predicted compute-optimal model, *Chinchilla*, that uses the same compute budget as *Gopher* but with 70B parameters and 4× more more data. *Chinchilla* uniformly and significantly outperforms *Gopher* (280B), GPT-3 (175B), Jurassic-1 (178B), and Megatron-Turing NLG (530B) on a large range of downstream evaluation tasks. This also means that *Chinchilla* uses substantially less compute for fine-tuning and inference, greatly facilitating downstream usage. As a highlight, *Chinchilla* reaches a state-of-the-art average accuracy of 67.5% on the MMLU benchmark, greater than a 7% improvement over *Gopher*.

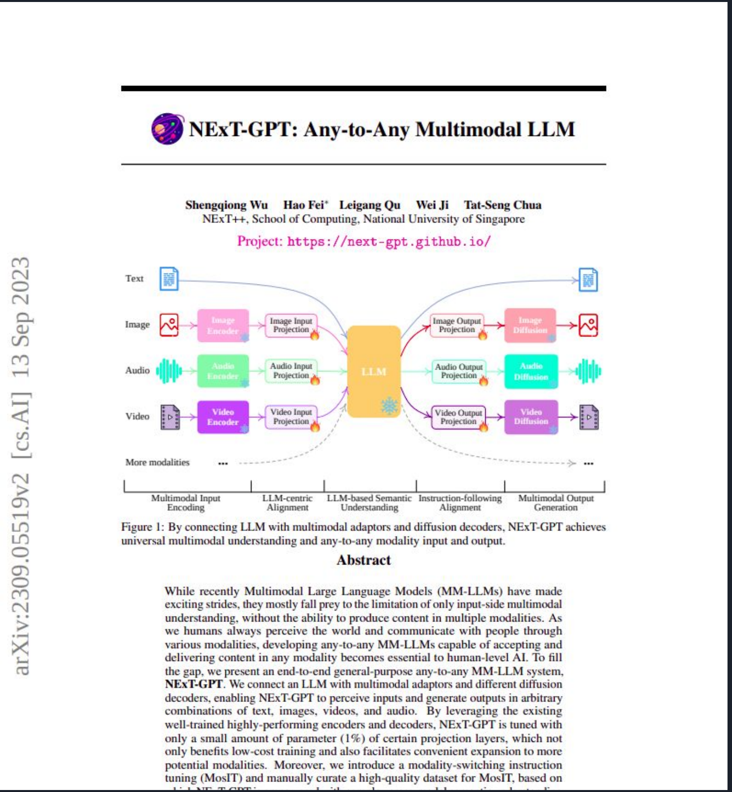
arXiv:2203.15556v1 [cs.CL] 29 Mar 2022

1. Introduction

Recently a series of *Large Language Models* (LLMs) have been introduced (Brown et al., 2020; Lieber et al., 2021; Rae et al., 2021; Smith et al., 2022; Thoppilan et al., 2022), with the largest dense language models now having over 500 billion parameters. These large autoregressive transformers (Vaswani et al., 2017) have demonstrated impressive performance on many tasks using a variety of evaluation protocols such as zero-shot, few-shot, and fine-tuning.

The compute and energy cost for training large language models is substantial (Rae et al., 2021; Thoppilan et al., 2022) and rises with increasing model size. In practice, the allocated training compute budget is often known in advance: how many accelerators are available and for how long we want to use them. Since it is typically only feasible to train these large models once, accurately estimating the best model hyperparameters for a given compute budget is critical (Tay et al., 2021).

Kaplan et al. (2020) showed that there is a power law relationship between the number of parameters in an autoregressive language model (LM) and its performance. As a result, the field has been training larger and larger models, expecting performance improvements. One notable conclusion in Kaplan et al. (2020) is that large models should not be trained to their lowest possible loss to be compute optimal. Whilst we reach the same conclusion, we estimate that large models should be



NExT-GPT: Any-to-Any Multimodal LLM

Shengqiong Wu   Hao Fei*   Leigang Qu   Wei Ji   Tat-Seng Chua
NExT++, School of Computing, National University of Singapore

Project: https://next-gpt.github.io/

Figure 1: By connecting LLM with multimodal adaptors and diffusion decoders, NExT-GPT achieves universal multimodal understanding and any-to-any modality input and output.

arXiv:2309.05519v2 [cs.AI] 13 Sep 2023

Abstract

While recently Multimodal Large Language Models (MM-LLMs) have made exciting strides, they mostly fall prey to the limitation of only input-side multimodal understanding, without the ability to produce content in multiple modalities. As we humans always perceive the world and communicate with people through various modalities, developing any-to-any MM-LLMs capable of accepting and delivering content in any modality becomes essential to human-level AI. To fill the gap, we present an end-to-end general-purpose any-to-any MM-LLM system, **NExT-GPT**. We connect an LLM with multimodal adaptors and different diffusion decoders, enabling NExT-GPT to perceive inputs and generate outputs in arbitrary combinations of text, images, videos, and audio. By leveraging the existing well-trained highly-performing encoders and decoders, NExT-GPT is tuned with only a small amount of parameter (1%) of certain projection layers, which not only benefits low-cost training and also facilitates convenient expansion to more potential modalities. Moreover, we introduce a modality-switching instruction tuning (MosIT) and manually curate a high-quality dataset for MosIT, based on

Getting Started

Research Projects

Choose to explore ideas that seem likely to succeed, are intriguing, or have the potential to lead to something new, or that contradict received wisdom.

It isn't possible to know whether the work is novel or will lead to valuable results.

The final outcome is an objective scientific report, but curiosity and guesswork are what establish research directions.

Research Projects

Establish answers to two key questions:

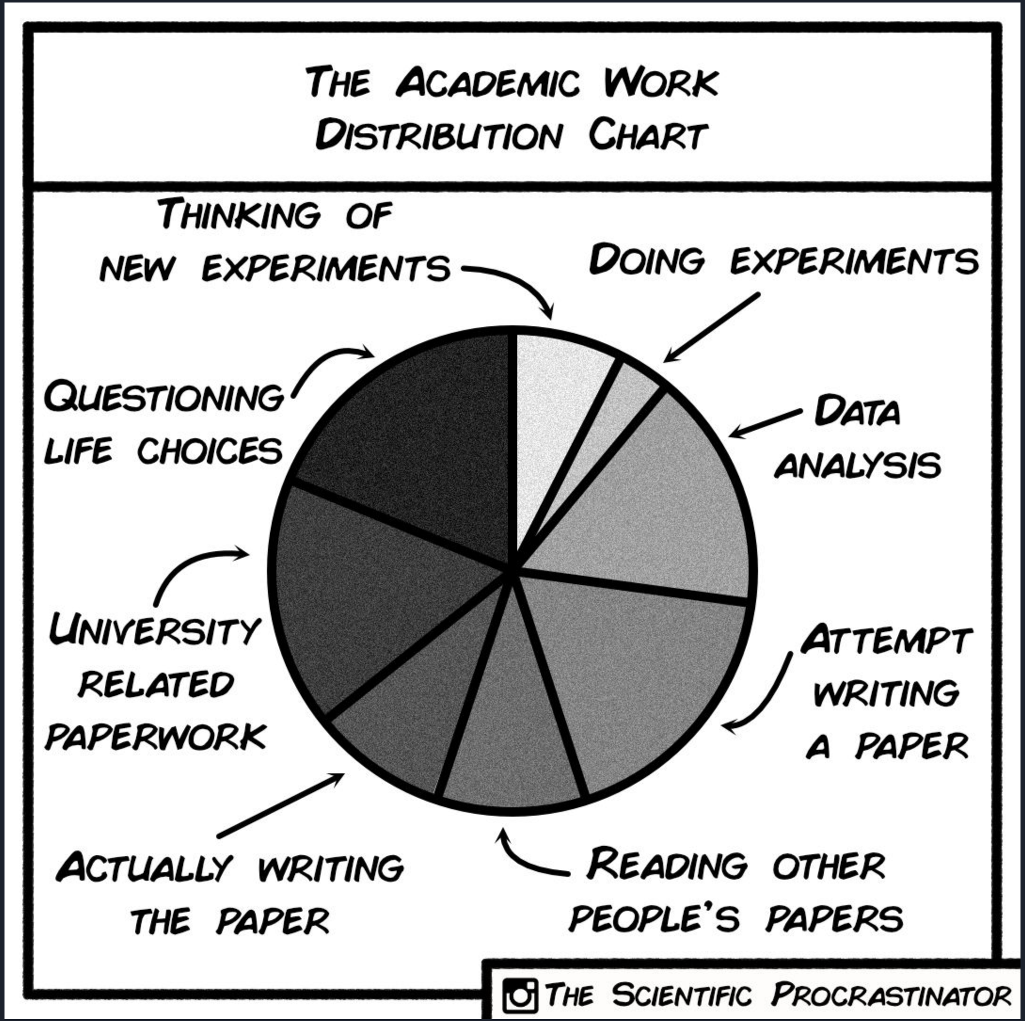
- First, what is the broad problem to be investigated?
- Second, what are the specific initial activities to undertake and outcomes to pursue?

Research Projects

Most research is to some extent incremental: it improves, repairs, extends, varies, or replaces work done by others.

The issue is the magnitude of the increment: a change to the fields in a network packet to save a couple of bits is unlikely to be worth investigating.

There needs to be challenge and the possibility of unexpected discovery for research to be interesting.


THE ACADEMIC WORK DISTRIBUTION CHART
THINKING OF NEW EXPERIMENTS
DOING EXPERIMENTS
DATA ANALYSIS
ATTEMPT WRITING A PAPER
READING OTHER PEOPLE'S PAPERS
ACTUALLY WRITING THE PAPER
UNIVERSITY RELATED PAPERWORK
QUESTIONING LIFE CHOICES
THE SCIENTIFIC PROCRASTINATOR

Reading Papers

Reading papers

Research is not primarily about running experiments, developing theory, or doing analysis.

Importance of developing an understanding: researchers do their best work after they have been in a field for five years or more.

To acquire this understanding, you need to become an effective reader of research papers.

Reading papers

It is important to become an effective reader:

- Skim through it to identify the extent to which it is relevant—only read it thoroughly if there is likely tobe value in doing so.
- Read through the abstract, results, and conclusion. Then decide if the rest is worth reading.
- Make the effort to properly understand the details, but always beware of details that may be wrong, or garbled.

Reading papers

Expect to have a range of modes of reading:

- Browsing to find papers and get an overview of activity and to understand the main outcomes in a body of work.
- Background reading of texts and popular science magazines.
- Thorough, focused reading of key or complex papers that stretch your abilities or the limits of your understanding.


Do my own research
Read all these important papers

Structure of a paper

Structure of a paper

Scientific papers follow a standard structure that allows readers to quickly discover the main results.

Many readers accept or reject conclusions based on a quick scan.

A well-structured write-up has important statements as near the beginning as possible.

Structure of a paper

You need to:

- Describe the work in the context of accepted scientific knowledge.
- State the idea that is being investigated, often as a theory or hypothesis.
- Explain what is new about the idea, or what contribution the paper is making.
- Justify the theory, by methods such as proof or experiment.