

Exploring Large Language Models

ChatGPT - What's the big deal?

Nils Vreman

December 13-15, 2023

Presentation Structure

- ▶ Every slide has been generated using ChatGPT
- ▶ Every image has been generated using DALL-E (one shot)
 - ▶ DALL-E is a modern text-to-image systems
 - ▶ Captions are summaries of the query used to generate the image
- ▶ If anything has been changed, it's marked in red

What is AI?

- ▶ Convolutional Neural Networks
 - ▶ Generative Adversarial Networks
 - ▶ Deep Neural Networks
 - ▶ Recurrent Neural Networks
 - ▶ Reinforcement Learning

▶ |

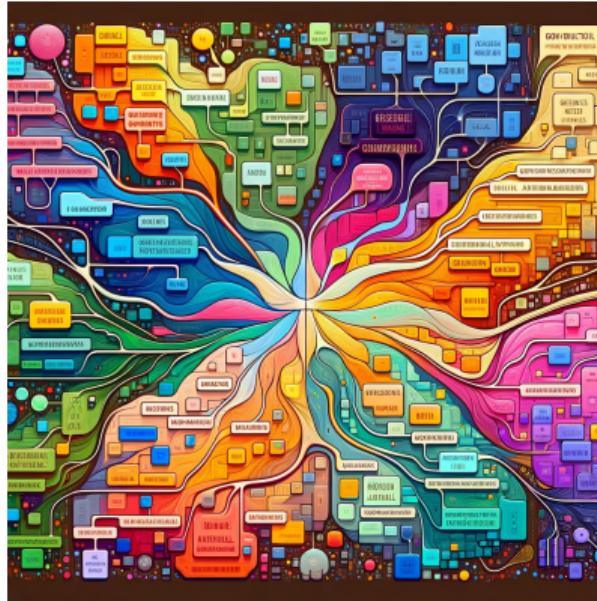


Figure: AI topology

What is Natural Language Processing?

Natural Language Processing (NLP) is a branch of AI that focuses on the interaction between computers and humans through natural language.

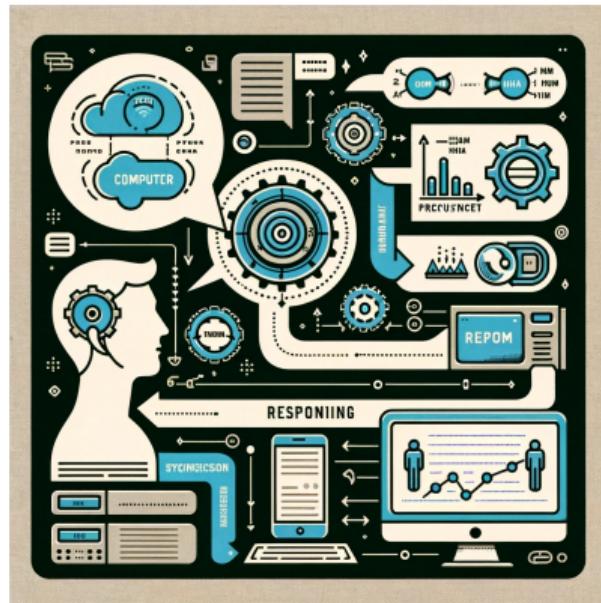


Figure: Infographic showing human ↔ computer interaction

Introduction to Language Models

Language models are AI models that understand, interpret, generate, and respond to human language.

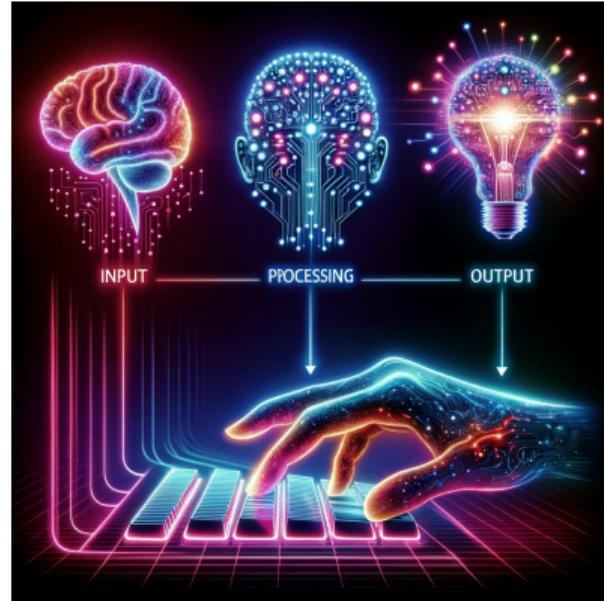


Figure: Visually stunning illustration of NLP

Rise of Large Language Models

Large Language Models represent a significant leap in AI's ability to process and generate human language, thanks to their scale and complexity.

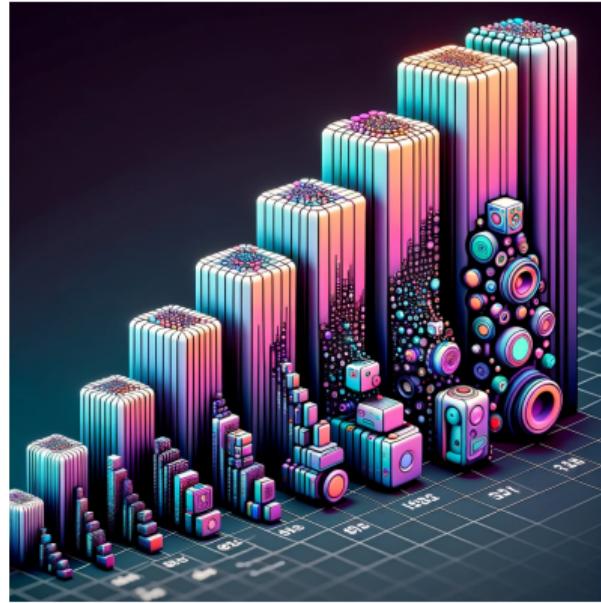


Figure: Comparison graphic of LLM against earlier models

Training Large Language Models

- ▶ Reinforcement Learning involves training models to make sequences of decisions by rewarding desired behaviors.
- ▶ In LLMs, it can be used for fine-tuning responses based on feedback.
- ▶ This in connection with human feedback is what makes ChatGPT so powerful

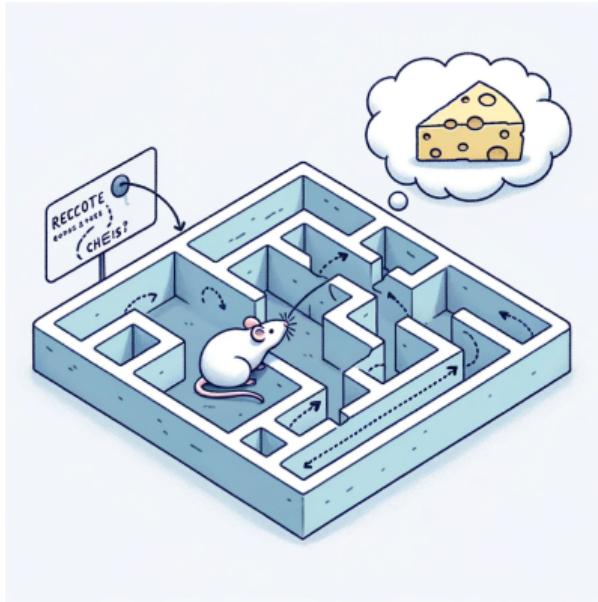


Figure: Reinforcement learning visualised as a mouse in a maze being rewarded for finding the middle

ChatGPT is Not Alone...

- ▶ GPT (OpenAI/[Microsoft](#))
- ▶ PaLM (Google)
- ▶ LLaMA (Meta)
- ▶ Gemini ([DeepMind/Google](#))
- ▶ Alpha Code (Google)
- ▶ Grok (X)

These models represent the cutting edge in natural language understanding and generation.

ChatGPT is Not Alone...

	Gemini Ultra	Gemini Pro	GPT-4	GPT-3.5	PaLM 2-L	Claude 2	Inflection-2	Grok 1	LLAMA-2
MMLU Multiple-choice questions in 57 subjects (professional & academic) (Hendrycks et al., 2021a)	90.04% CoT@32* 83.7% 5-shot	79.13% CoT@8* 71.8% 5-shot	87.29% CoT@32 (via API**) 86.4% 5-shot (reported)	70% 5-shot	78.4% 5-shot	78.5% 5-shot CoT	79.6% 5-shot	73.0% 5-shot	68.0%***
GSM8K Grade-school math (Colbe et al., 2021)	94.4% Maj1@32	86.5% Maj1@32	92.0% SFT & 5-shot CoT	57.1% 5-shot	80.0% 5-shot	88.0% 0-shot	81.4% 8-shot	62.9% 8-shot	56.8% 5-shot
MATH Math problems across 5 difficulty levels & 7 subdisciplines (Hendrycks et al., 2021b)	53.2% 4-shot	32.6% 4-shot	52.9% 4-shot (via API**) 50.3% (Zheng et al., 2023)	34.1% 4-shot (via API**)	34.4% 4-shot	—	34.8% 8-shot	23.9% 4-shot	13.5% 4-shot
BIG-Bench-Hard Subset of hard BIG-bench tasks written as CoT problems (Srivastava et al., 2022)	83.6% 3-shot	75.0% 3-shot	83.1% 3-shot (via API**)	66.6% 3-shot (via API**)	77.7% 3-shot	—	—	—	51.2% 3-shot
HumanEval Python coding tasks (Chen et al., 2021)	74.4% 0-shot (IT)	67.7% 0-shot (IT)	67.0% 0-shot (reported)	48.1% 0-shot	—	70.0% 0-shot	44.5% 0-shot	63.2% 0-shot	29.9% 0-shot
Natural2Code Python code generation. (New held-out set with no leakage on web)	74.9% 0-shot	69.6% 0-shot	73.9% 0-shot (via API**)	62.3% 0-shot (via API**)	—	—	—	—	—
DROP Reading comprehension & arithmetic. (metric: F1-score) (Dua et al., 2019)	82.4 Variable shots	74.1 Variable shots	80.9 3-shot (reported)	64.1 3-shot	82.0 Variable shots	—	—	—	—
HellaSwag (validation set) Common-sense multiple choice questions (Zellers et al., 2019)	87.8% 10-shot	84.7% 10-shot	95.3% 10-shot (reported)	85.5% 10-shot	86.8% 10-shot	—	89.0% 10-shot	—	80.0%***
WMT23 Machine translation (metric: BLEURT) (Tom et al., 2023)	74.4 1-shot (IT)	71.7 1-shot	73.8 1-shot (via API**)	—	72.7 1-shot	—	—	—	—

Figure: Comparison of LLMs (not generated by DALL-E)

ChatGPT for Developers

Leveraging ChatGPT in software development:

- ▶ Automating code generation, **testing**, and debugging.
- ▶ Integrating natural language interfaces into applications.
- ▶ Enhancing software documentation and support systems.
- ▶ Lowering the threshold for new developers to join a project

Ideal for streamlining development workflows and improving productivity.



Figure: Developer working with AI assistance

ChatGPT for Management

Empowering management with ChatGPT:

- ▶ Enhancing decision-making with data-driven insights.
- ▶ Streamlining report generation, **translation**, and documentation.
- ▶ Providing interactive tools for strategy planning and analysis.

A valuable asset for efficient and informed managerial decisions.



Figure: Board room where ChatGPT is assisting

ChatGPT for Economists

ChatGPT's role in economics:

- ▶ Conducting in-depth market analysis and forecasting.
- ▶ Assisting in economic research and data interpretation.
- ▶ Predicting trends based on large datasets.

A powerful tool for data-driven economic insights and research.



Figure: Economist creating market forecasts using ChatGPT

ChatGPT in Personal Life: Learning and Productivity

Enhancing daily life with ChatGPT:

- ▶ Learning new skills and languages.
- ▶ Personalized productivity and organization tools.
- ▶ Customized educational support and tutoring.

A personal assistant to foster continuous learning and efficiency.



Figure: People using AI for educational purposes

ChatGPT in Personal Life: Entertainment and Creativity

ChatGPT for leisure and creativity:

- ▶ Generating creative writing, music, and art ideas.
- ▶ Interactive entertainment, like games and storytelling.
- ▶ Planning travel and leisure activities.

A source of inspiration and fun in everyday life.



Figure: Family creating telling AI generated stories together

The Rapid Advancements of LLMs

- ChatGPT was released 22nd November 2022
- Significant improvements in accuracy and understanding.
- Expanded applications in various sectors.

A year of rapid growth setting the stage for future developments.

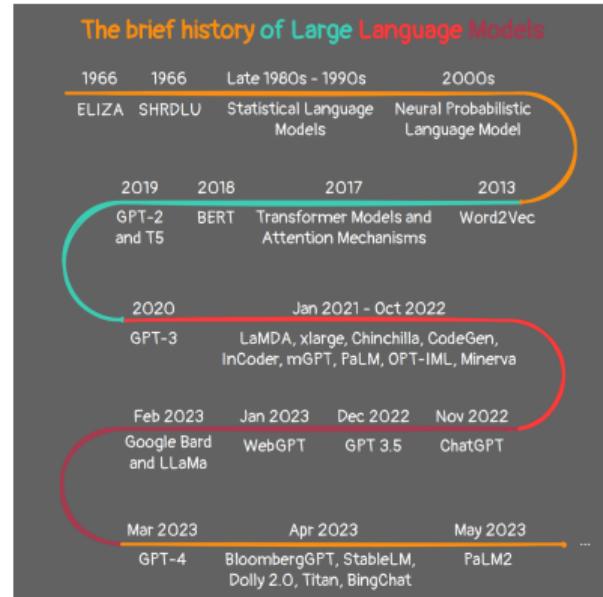


Figure: Borrowed from a Level Up Coding blog post by Armin Norouzi, Ph.D

Predictions for the Future of LLMs

Taking into account that the large language models of today are built on extremely simple principles:

- ▶ Anticipated advancements in AI capabilities and autonomy.
- ▶ Potential for transformative impacts across industries.
- ▶ Ethical and societal considerations.

Envisioning a future shaped by continuous AI innovation.



Figure: Conceptual illustration of the future of LLMs

The Importance of Awareness

some text

Adapting to the Future of LLMs

The importance of embracing AI advancements:

- ▶ Staying ahead in a rapidly evolving technological landscape.
- ▶ Leveraging AI for competitive advantage and innovation.
- ▶ Preparing for ethical and practical challenges.

Not joining in is not just a missed opportunity but a strategic oversight.



Figure: Evolving technological landscape with metaphors

Transformers - The Core of LLMs

- ▶ Transformers are a type of model architecture that's particularly effective for processing sequential data like text.
 - ▶ The use of "**self-attention**" allows transformers to learn the relationships between words in a sentence.

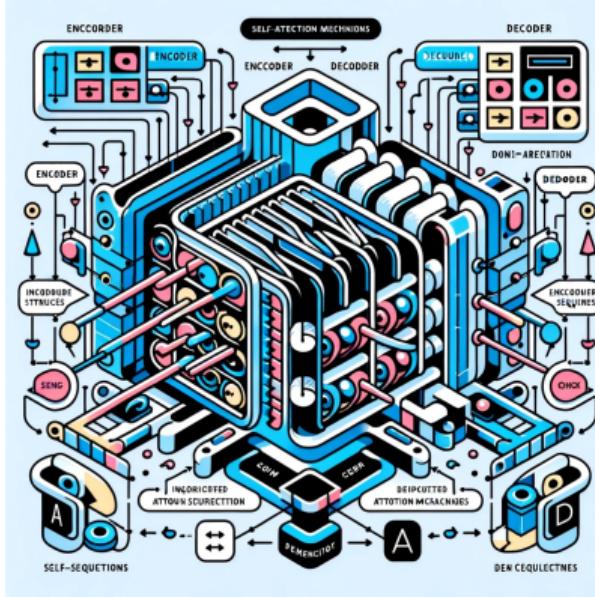


Figure: A simple representation of the encoder-decoder architecture of a transformer (This one physically hurt to allow)

Tokenization - Turning Text into Data

Tokenization is the process of converting text into smaller units (tokens) that can be easily processed by LLMs.

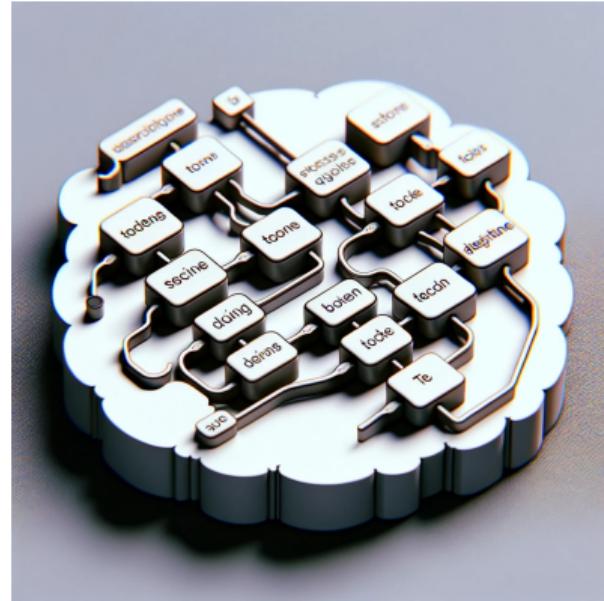


Figure: A sentence being broken down into tokens (words or subwords), maybe as a flow-diagram

Embeddings - Understanding Word Meaning

Embeddings are a way to represent tokens as vectors in a high-dimensional space, capturing their meaning and relationships.

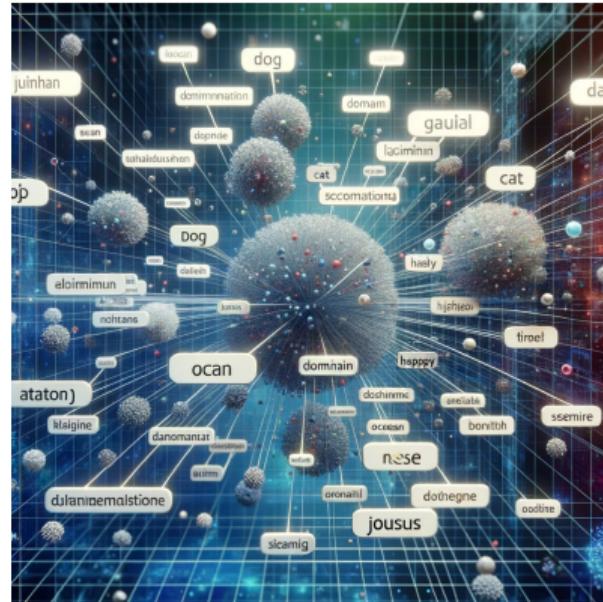


Figure: A 3D visualisation showing similar words being close together

Lorem Ipsum

Text