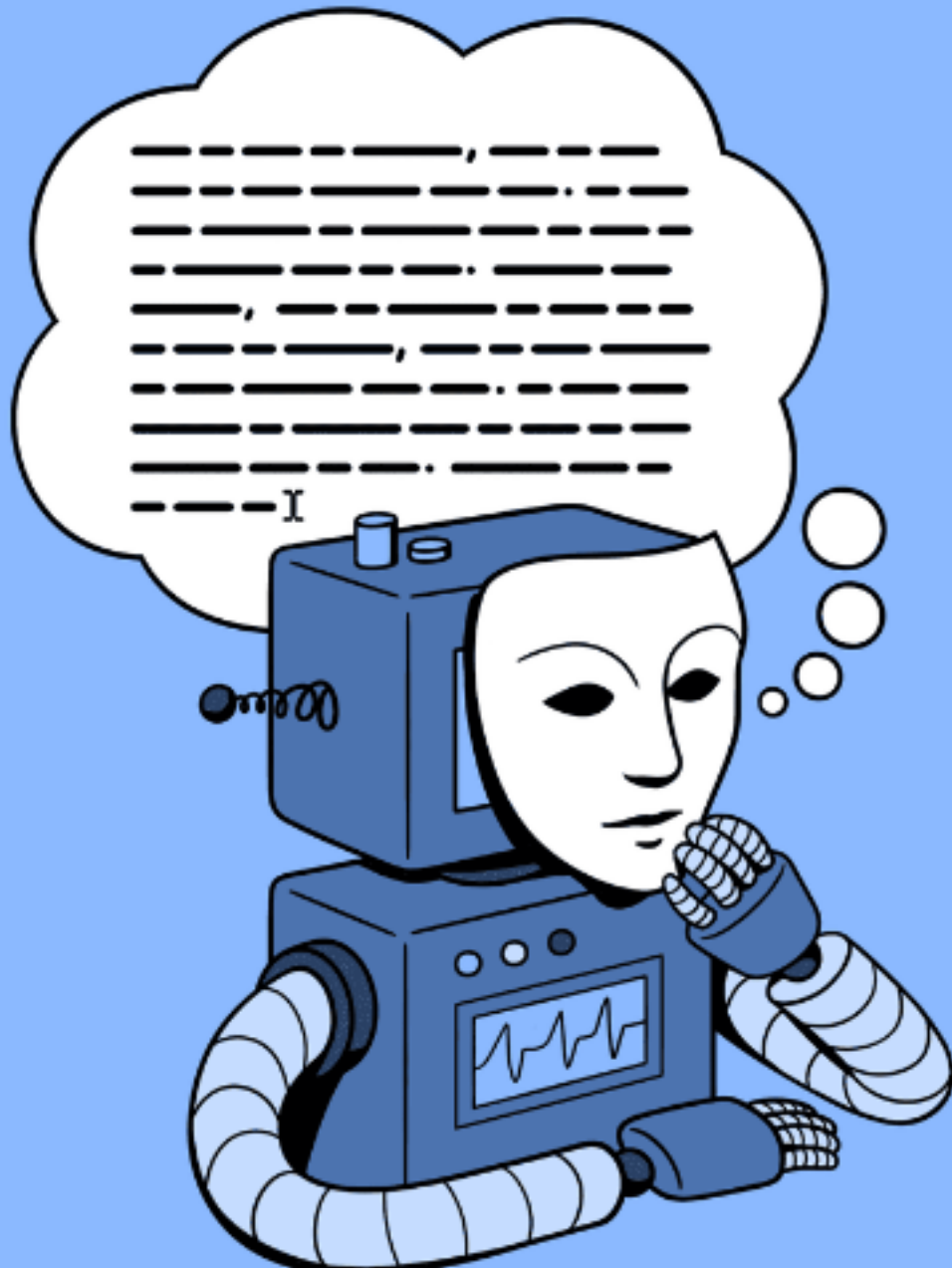




Getting Started with

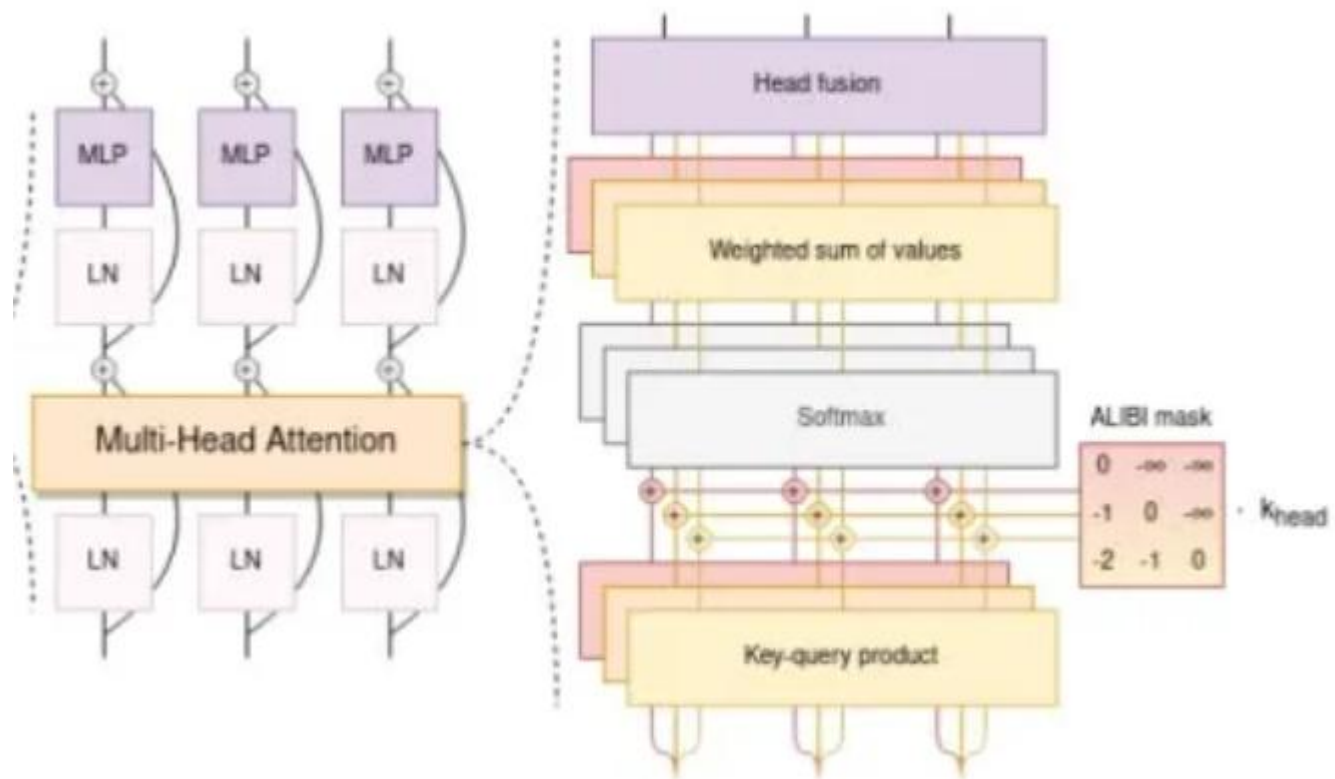
LLM



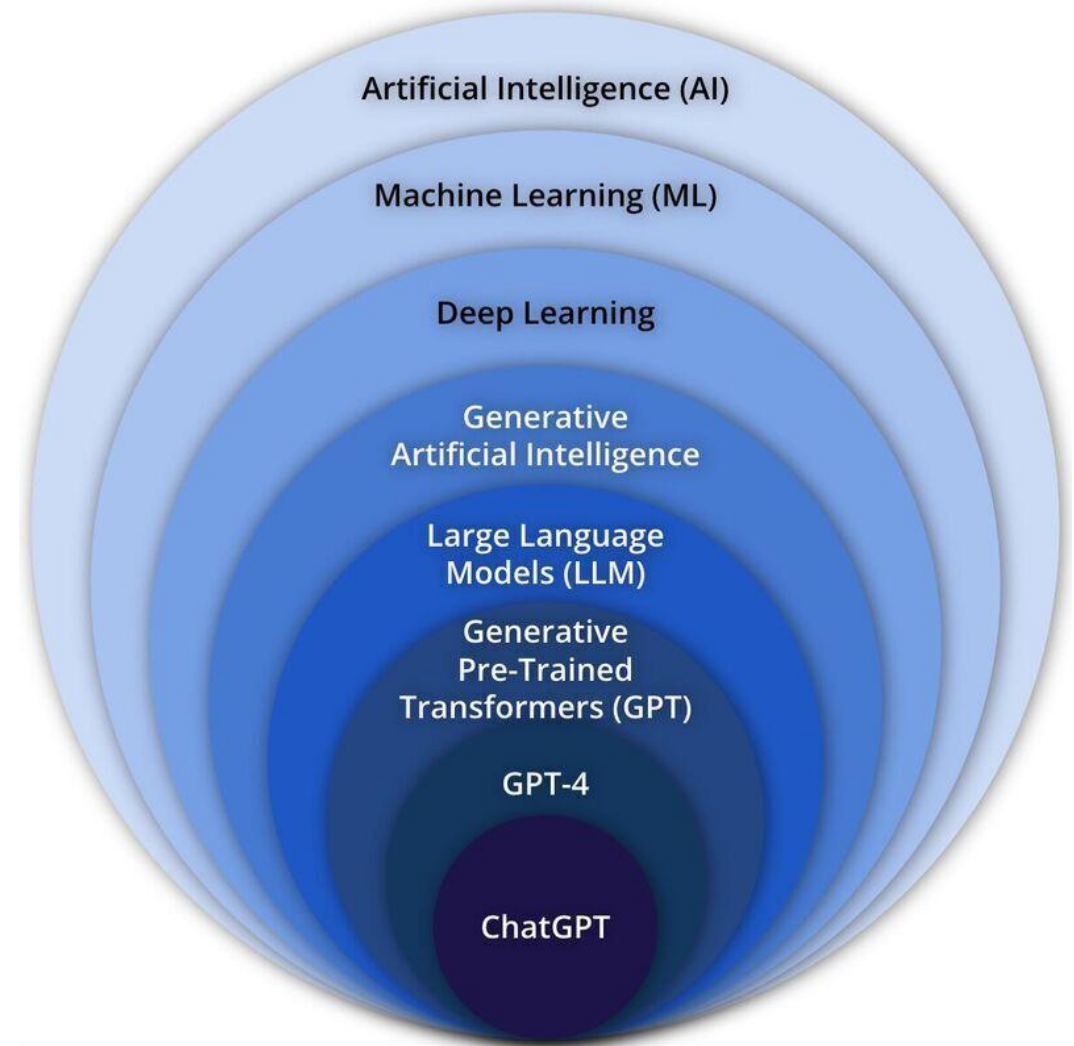
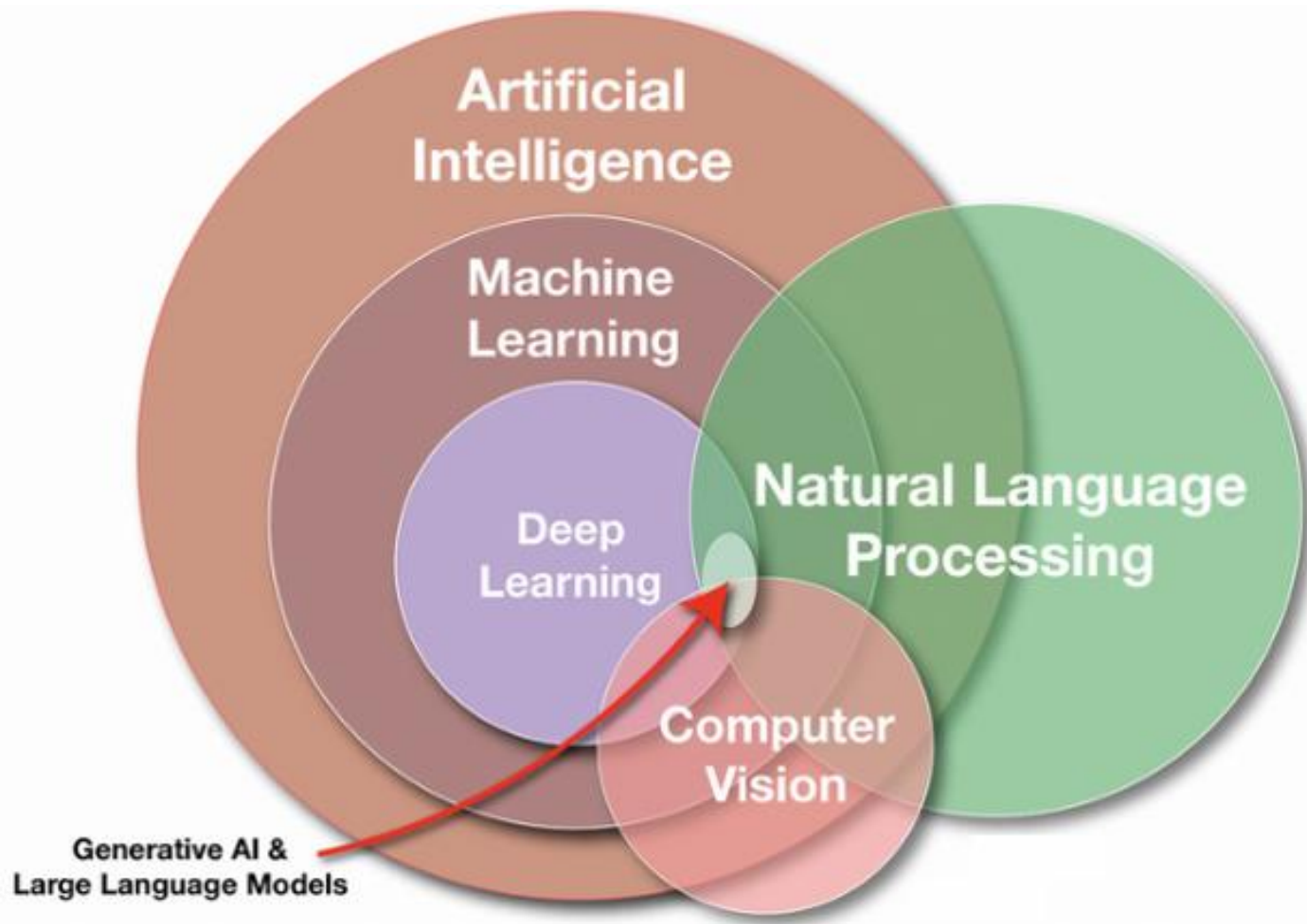
Large Language Model (LLM)

['lärj 'laŋ-gwij 'mä-dəl]

A deep learning algorithm that's equipped to summarize, translate, predict, and generate human-sounding text to convey ideas and concepts.



Placing Generative AI and LLMs in the bigger picture



Generative AI use cases across different data modalities

The Generative AI Application Landscape



APPLICATION LAYER	Marketing (content)						
	Sales (email)	Code generation	Image generation				Gaming
	Support (chat / email)	Code documentation	Consumer / Social				RPA
	General writing	Text to SQL	Media / Advertising				Music
	Note taking	Web app builders	Design	Voice Synthesis	Video editing / generation	3D models / scenes	Audio
	Other						Biology & chemistry
	TEXT	CODE	IMAGE	SPEECH	VIDEO	3D	OTHER
MODEL LAYER	OpenAI GPT-3	OpenAI GPT-3	OpenAI Dall-E 2	OpenAI	Microsoft X-CLIP	DreamFusion	TBD
	DeepMind Gopher	Tabnine	Stable Diffusion		Meta Make-A-Video	NVIDIA GET3D	
	Facebook OPT	Stability.ai	Craiyon			MDM	
	Hugging Face Bloom						
	Cohere						
	Anthropic						
	AI2						
	Alibaba, Yandex, etc.						

Discriminative vs Generative NLP Models:

Different Use-Cases

Discriminative and generative machine learning language models have different strengths and weaknesses, so they are used for different tasks.

Discriminative language models are better at tasks that require understanding the relationship between words and their meaning. For example, they can be used for:

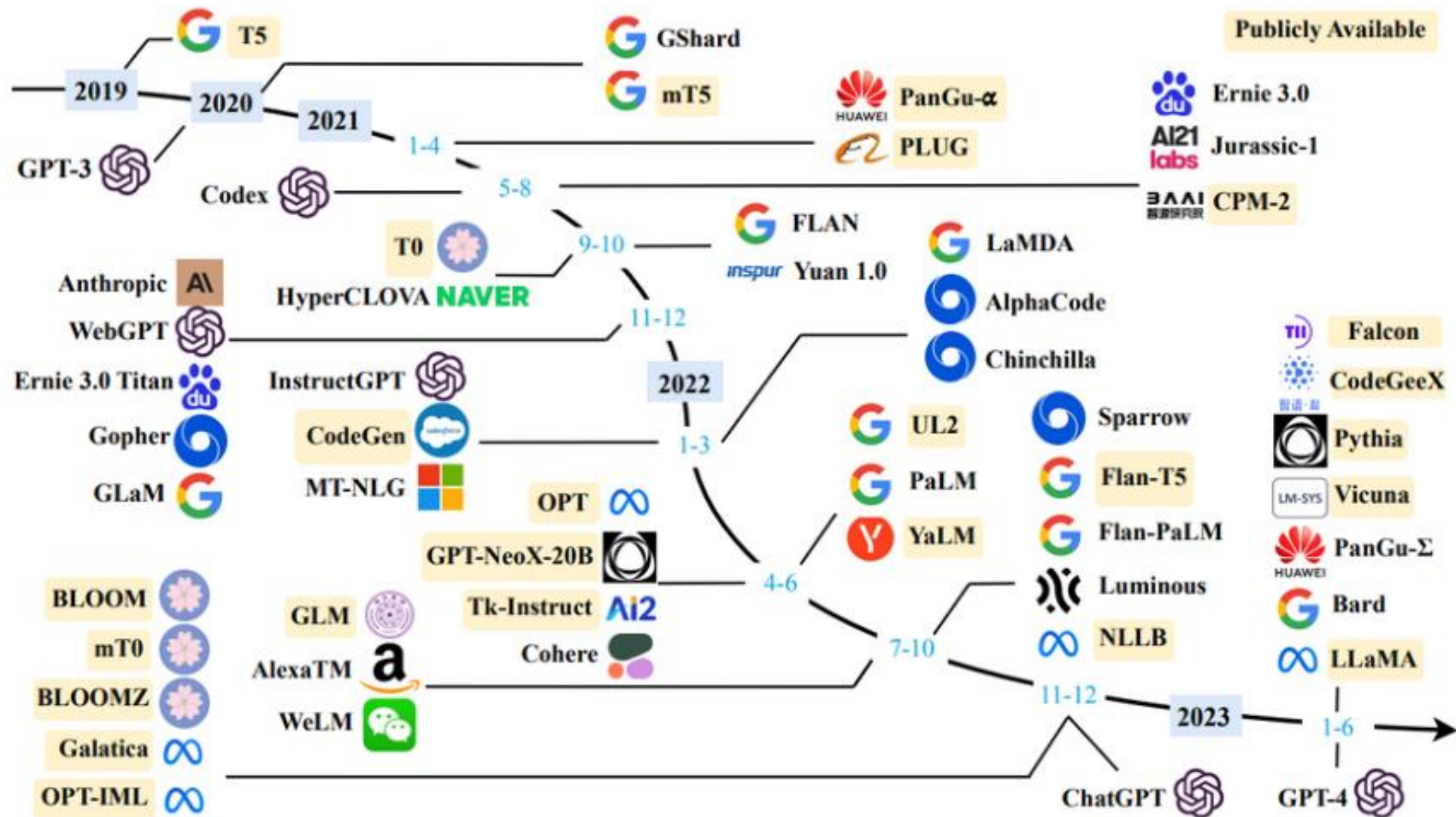
- **Text classification:** Categorizing text into different classes, such as news articles, product reviews, or spam.
- **Named entity recognition:** Identifying named entities in text, such as people, places, and organizations.
- **Sentiment analysis:** Identifying the sentiment of text, such as whether it is positive, negative, or neutral.

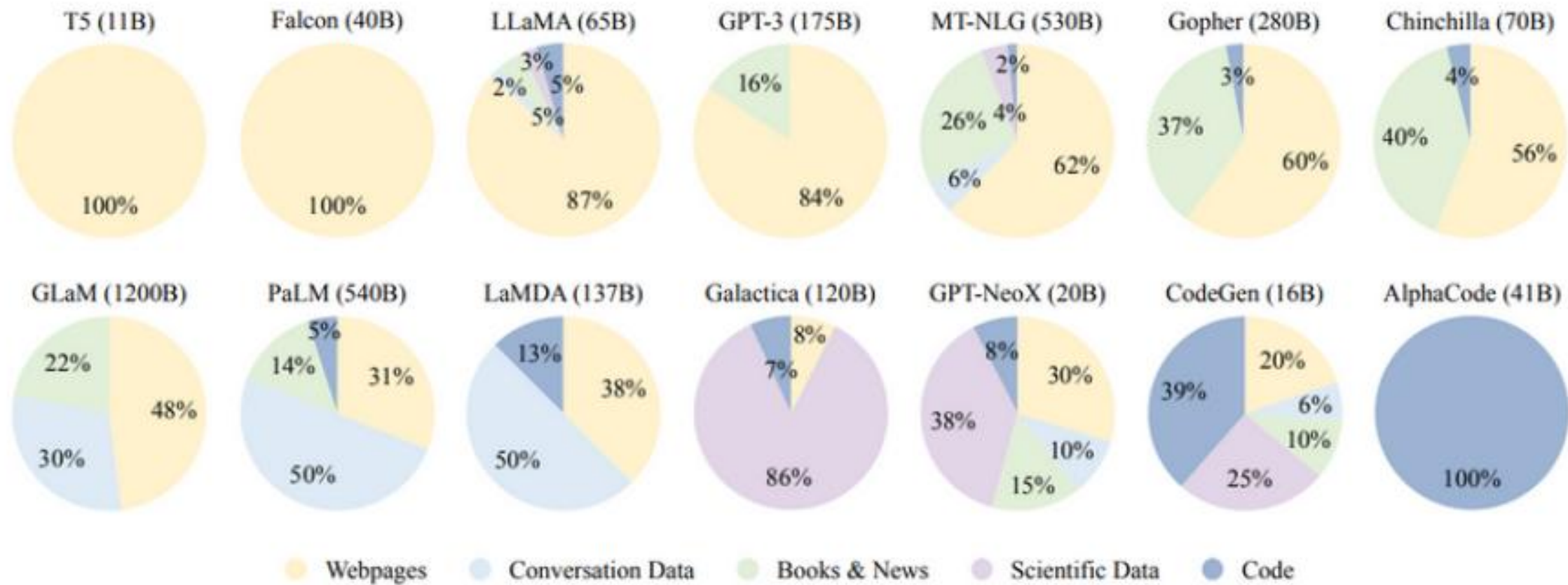
Generative language models are better at tasks that require creating new text. For example, they can be used for:

- **Text summarization:** Generating a shorter version of a text that captures the main points.
- **Machine translation:** Translating text from one language to another.
- **Text generation:** Generating new text, such as poems, code, or scripts.

Here are some specific examples of discriminative and generative language models:

- **Discriminative language models:** Logistic regression, support vector machines, conditional random fields etc
- **Generative language models:** Naive Bayes, Bayesian networks, hidden Markov models, etc.





Ratios of various data sources in the pre-training data for existing LLMs, arXiv:2303.18223.

Welcome to AI's new inflection point

ChatGPT has woken up the world to the transformative potential of artificial intelligence (AI), capturing global attention and sparking a wave of creativity rarely seen before. Its ability to mimic human dialogue and decision-making has given us AI's first true inflection point in public adoption. Finally, everyone, everywhere can see the technology's true disruptive potential for themselves.

ChatGPT reached 100 million monthly active users just two months after launch, making it the fastest-growing consumer application in history.¹

A foundation model is a generic term for large models with billions of parameters. With recent advances, companies can now build specialized image- and language-generating models on top of these foundation models. Large language models (LLMs) are both a type of generative AI and a type of foundation model.

The LLMs behind ChatGPT mark a significant turning point and milestone in artificial intelligence. Two things make LLMs game changing. First, they've cracked the code on language complexity. Now, for the first time, machines can learn language, context and intent and be independently generative and creative. Second, after being pre-trained on vast quantities of data (text, images or audio), these models can be adapted or fine-tuned for a wide range of tasks. This allows them to be reused or repurposed in many different ways.

Business leaders recognize the significance of this moment. They can see how LLMs and generative AI will fundamentally transform everything from business, to science, to society itself—unlocking new performance frontiers. The positive impact on human creativity and productivity will be massive. Consider that, across all industries, Accenture found 40% of all working hours can be impacted by LLMs like GPT-4. This is because language tasks account for 62% of the total time employees work, and 65% of that time can be transformed into more productive activity through augmentation and automation (see Figure 3).

How did we get here?

Milestones in the journey to generative AI



Machine learning: Analysis and prediction phase

The first decade of the 2000s marked the rapid advance of various machine learning techniques that could analyze massive amounts of online data to draw conclusions – or “learn” – from the results. Since then, companies have

viewed machine learning as an incredibly powerful field of AI for analyzing data, finding patterns, generating insights, making predictions and automating tasks at a pace and on a scale that was previously impossible.

Deep learning: Vision and speech phase

The 2010s produced advances in AI’s perception capabilities in the field of machine learning called deep learning. Breakthroughs in deep learning enable the computer vision

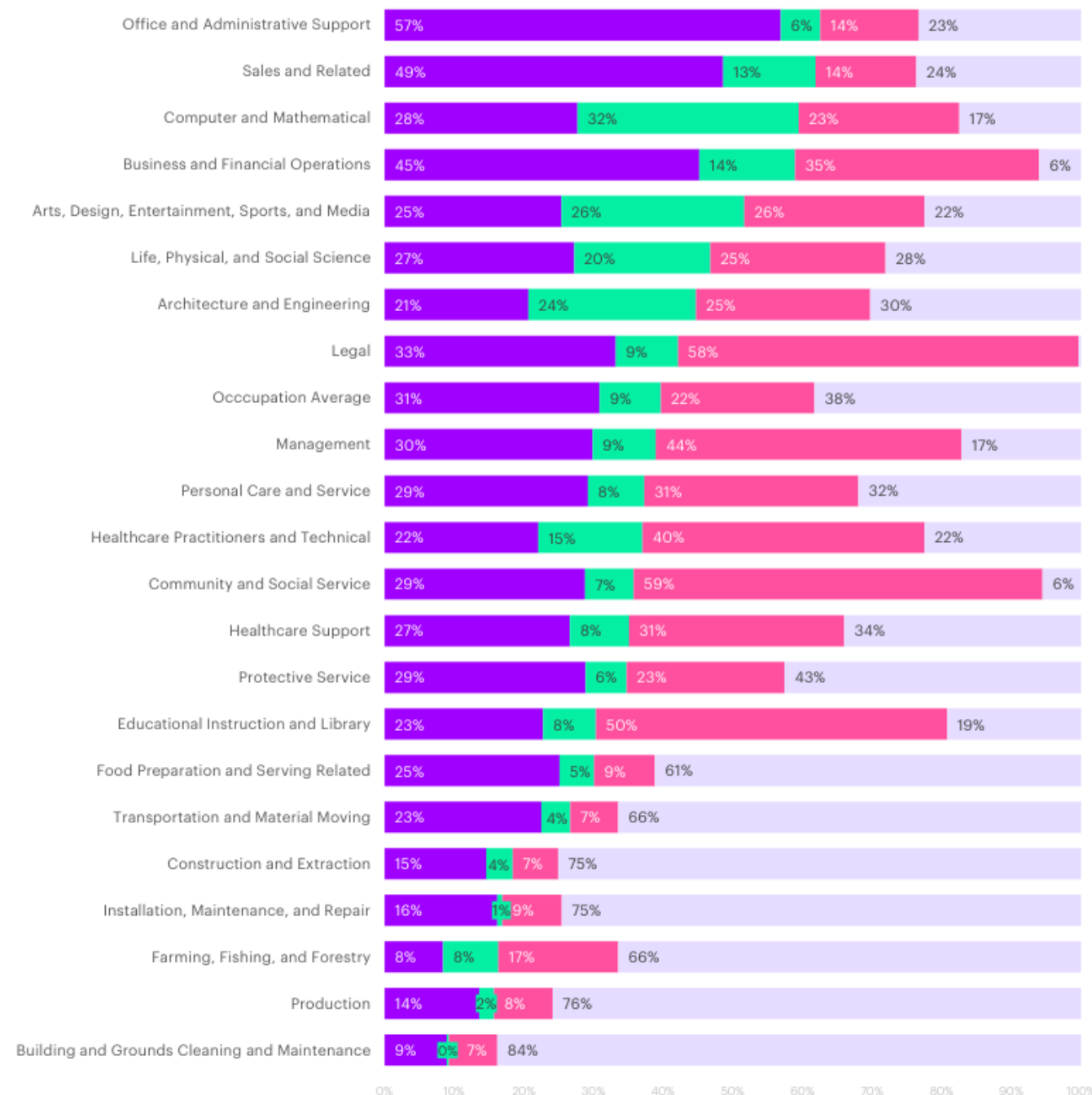
that search engines and self-driving cars use to classify and detect objects, as well as the voice recognition that allows popular AI speech assistants to respond to users in a natural way.

Generative AI: Enter the language-mastery phase

Building on exponential increases in the size and capabilities of deep learning models, the 2020s will be about language mastery. The GPT-4 language model, developed by OpenAI, marks the beginning of a new

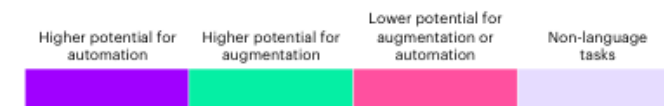
phase in the abilities of language-based AI applications. Models such as this will have far-reaching consequences for business, since language permeates everything an organization does day to day—its institutional knowledge, communication and processes.²

Figure 4: Generative AI will transform work across every job category



Work time distribution by major occupation and potential AI impact

Based on their employment levels in the US in 2021



In 5 out of 22 occupation groups, Generative AI can affect more than half of all hours worked

Source: Accenture Research based on analysis of Occupational Information Network (O*NET), US Dept. of Labor; US Bureau of Labor Statistics.

Notes: We manually identified 200 tasks related to language (out of 332 included in BLS), which were linked to industries using their share in each occupation and the occupations' employment level in each job category. Tasks with higher potential for automation can be transformed by LLMs with reduced involvement from a human worker. Tasks with higher potential for augmentation are those in which LLMs would need more involvement from human workers.

Here are some examples of how LLMs can be used:

- **Text summarization:** An LLM can be used to generate a shorter version of a text that captures the main points. For example, an LLM could be used to summarize a news article or a research paper.
- **Machine translation:** An LLM can be used to translate text from one language to another. For example, an LLM could be used to translate a website from English to Spanish.
- **Text generation:** An LLM can be used to generate new text, such as poems, code, or scripts. For example, an LLM could be used to generate a new poem or to write a new computer program.

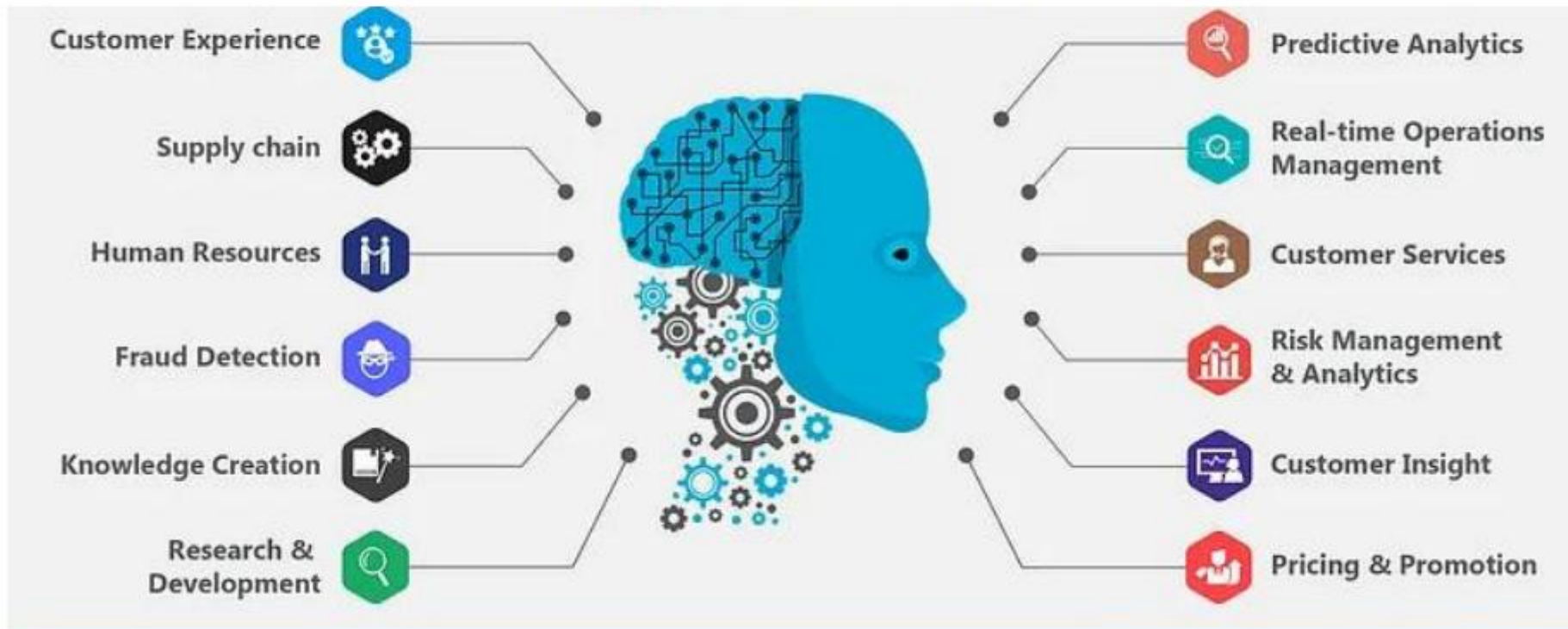
LLMs are still under development, but they have the potential to revolutionize the way we interact with computers. In the future, we can expect to see LLMs being used for a wider range of tasks, such as creating realistic chatbots and generating creative content.

Use Cases of LLMs across industries

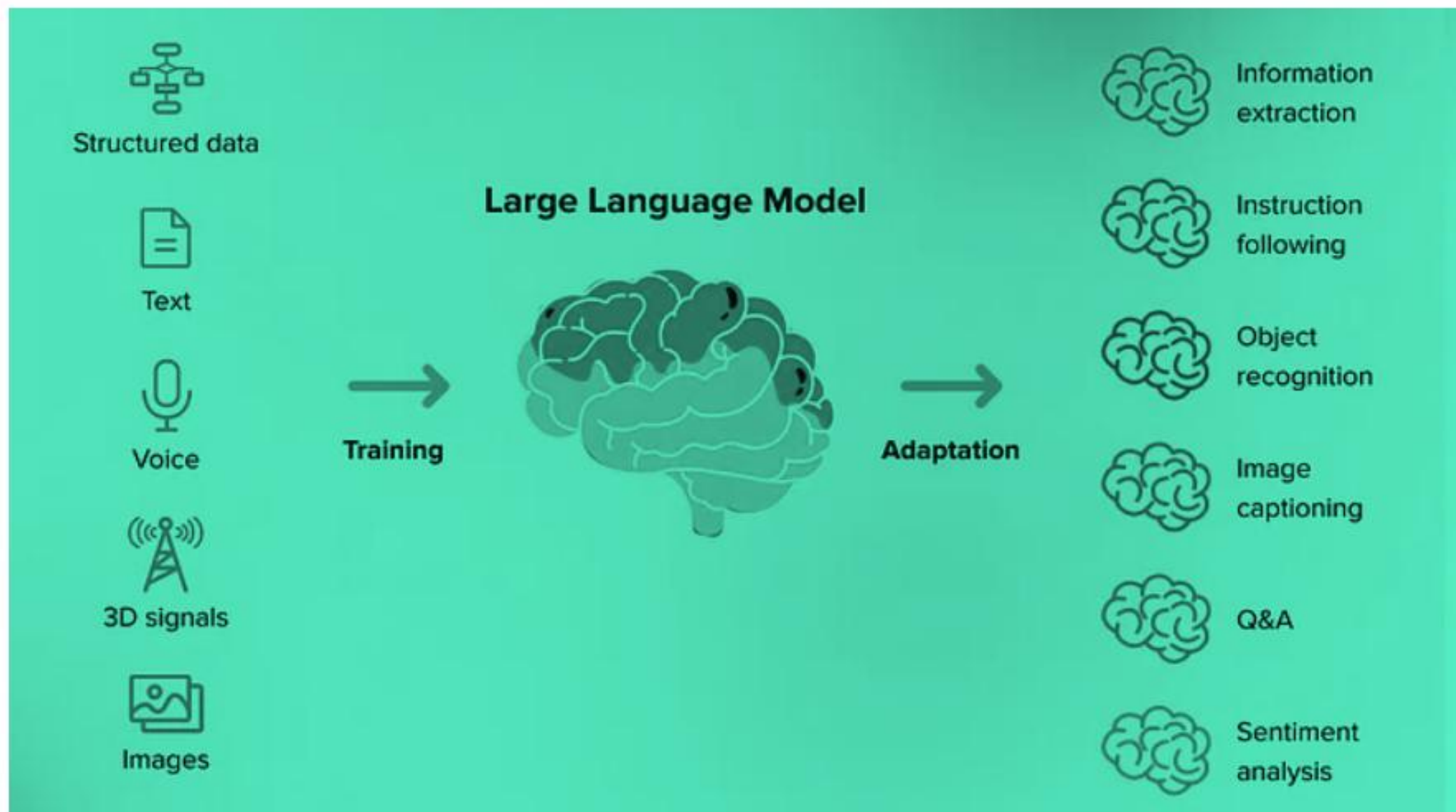
Large Language Models (LLMs) are a rapidly expanding field of research and development. They are being used in a variety of industries, including healthcare, retail, tech, and more. Here are some common use cases of LLMs:

1. **Text summarization:** LLMs can summarize blocks of text or multiple documents into a shorter version while retaining the most important information
2. **Text generation:** LLMs can generate text on any topic they are trained on. This can be used to create chatbots, virtual assistants, and conversational AI.
3. **Sentiment analysis:** LLMs can analyze text to determine the sentiment behind it. This can be used to gauge customer satisfaction or public opinion.
4. **Content creation:** LLMs can be used to create content such as articles, stories, and even poetry.
5. **Question answering:** LLMs can answer questions based on the context provided to them. This can be used in chatbots or virtual assistants to provide quick answers to common questions
6. **Clustering:** LLMs can group similar data points together based on their similarity in meaning or context.
7. **Classification:** LLMs can classify data points into different categories based on their content or meaning.

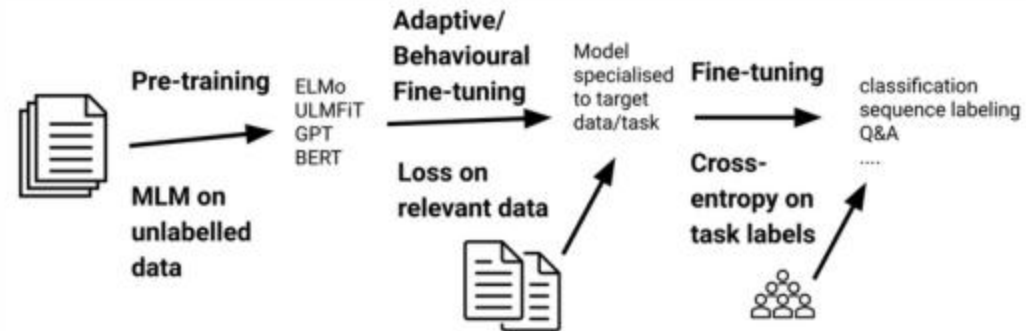
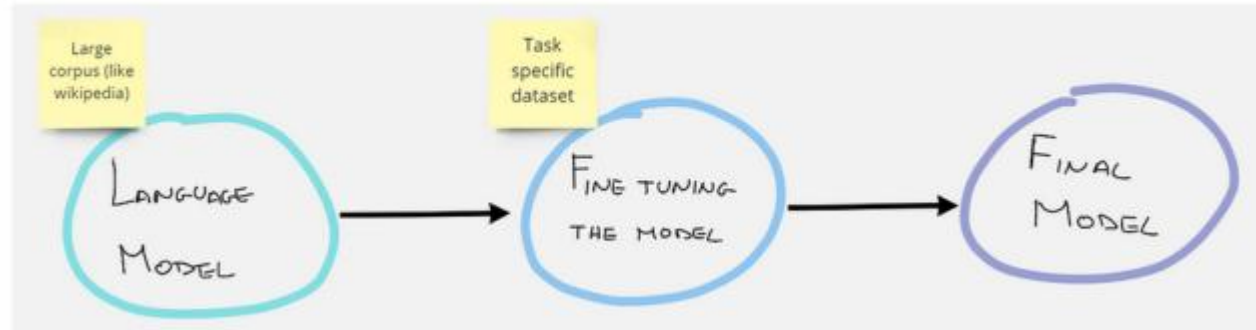
AI/ML Use-Cases

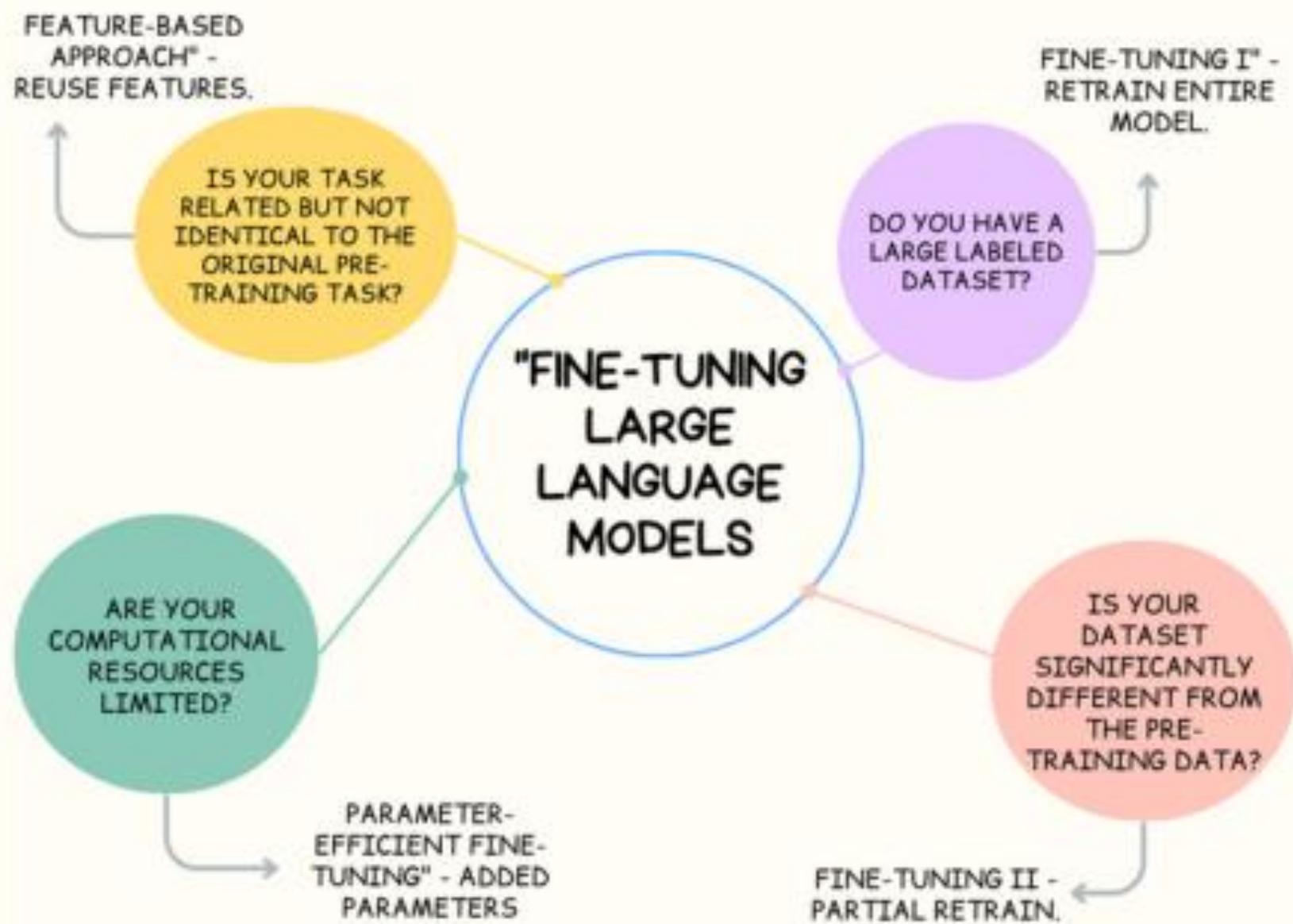


Adapting LLMs in enterprise



Fine-Tuning Large Language Models





GLUE Benchmark

The General Language Understanding Evaluation (GLUE) benchmark comprises nine diverse natural language understanding tasks, serving as a comprehensive yardstick for assessing various LM models. The GLUE benchmark proves effective in evaluating LMs designed for versatile applications.

Task-Specific Downstream Evaluation

Alternatively, it might be more pertinent to appraise an LM's efficacy based on its designated task. For instance, an LM tailored for text classification can be assessed through conventional classification metrics like precision, recall, and F1 score.

Perplexity

Perplexity serves as a statistical gauge of an LM's text prediction confidence. Lower perplexity values reflect adept test set prediction, while higher values denote inadequate prediction. This metric is pertinent for evaluating LMs designed for text generation and machine translation tasks.

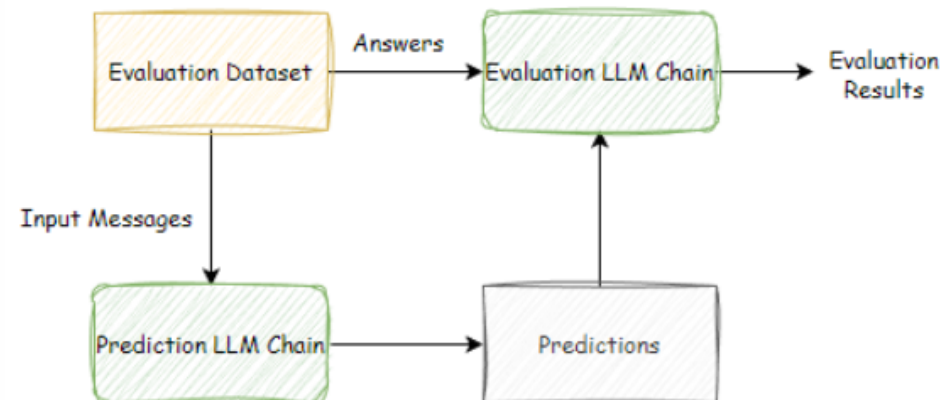
BLEU Score

The BLEU (bilingual evaluation understudy) score, ranging from 0 to 1, gauges the quality of machine translation compared to a reference translation. Higher BLEU scores indicate greater similarity to the reference text. This metric holds significance in evaluating LMs geared towards machine translation tasks.

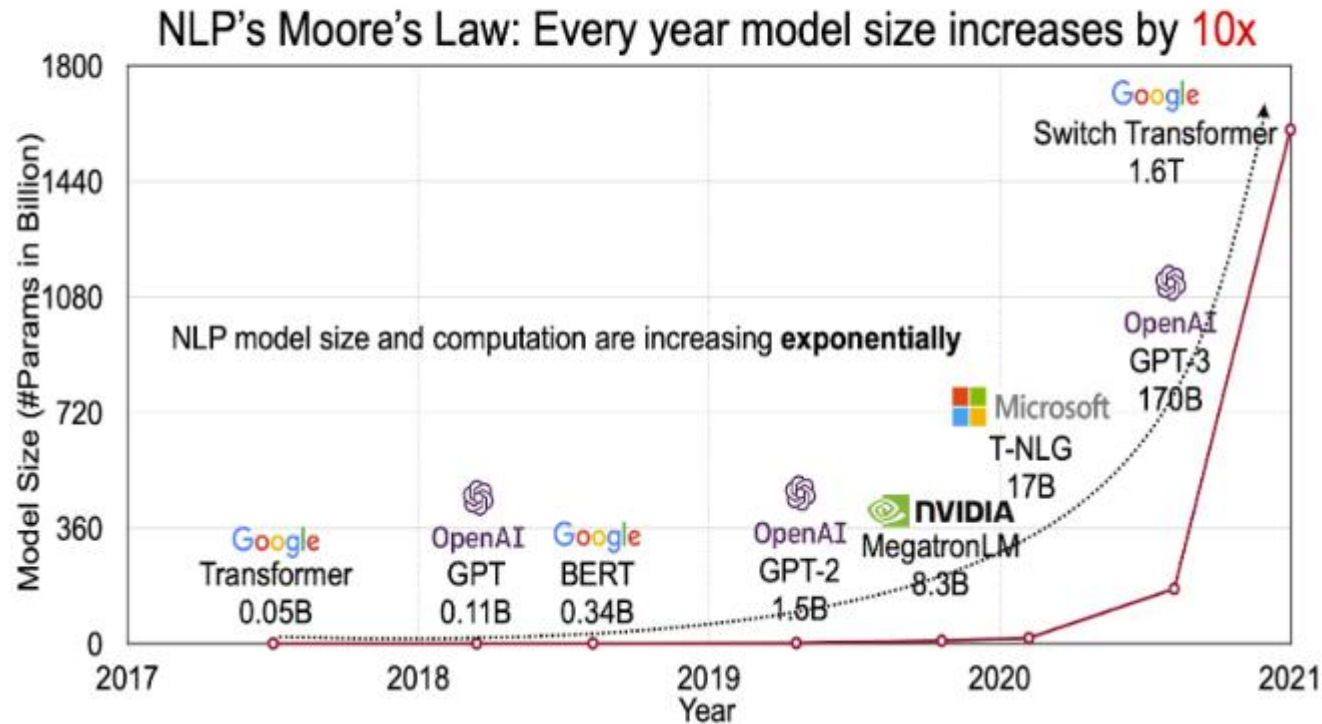
Human Evaluation

In conjunction with statistical and automated evaluation methodologies, human evaluators play a pivotal role in assessing LMs' attributes such as creativity, humor, and toxicity. Their insights offer valuable feedback on the quality of LM-generated content.

The selection of the most suitable evaluation approach for an LM hinges on its designated application. However, the aforementioned methodologies present a strong foundational framework for LM evaluation.



Not just hype: LLMs are here to stay



Key Industries taking advantage of LLMs in conversational technology

- **Virtual assistants:** LLMs are being used to create virtual assistants that can provide information and complete tasks for users. For example, Amazon's Alexa and Apple's Siri are both powered by LLMs.
- **Call centers:** LLMs are being used to create chatbots that can answer customer questions and resolve issues. This can help businesses to reduce the cost of providing customer support.
- **Education:** LLMs are being used to create educational tools that can help students learn new languages and concepts. For example, Google's AIY Project includes a voice-activated language learning app that uses an LLM to generate realistic and natural-sounding speech.

This is a pivotal moment. For several years, generative AI and foundation models have been quietly revolutionizing the way we think about machine intelligence. Now, thanks to ChatGPT, the whole world has woken up to the possibilities this creates.

While artificial general intelligence (AGI) remains a distant prospect, the speed of development continues to be breathtaking. We're at the start of an incredibly exciting era that will fundamentally transform the way information is accessed, content is created, customer needs are served, and businesses are run.

Embedded into the enterprise digital core, generative AI, LLMs, and foundation models will optimize tasks, augment human capabilities, and open up new avenues for growth. In the process, these technologies will create an entirely new language for enterprise reinvention.

Businesses are right to be optimistic about the potential of generative AI to radically change how work gets done and what services and products they can create. They also need to be realistic about the challenges that come with profoundly rethinking how the organization works, with implications for IT, organization, culture, and responsibility by design.

Companies need to invest as much in evolving operations and training people as they do in technology. Radically rethinking how work gets done, and helping people keep up with technology-driven change, will be two of the most important factors in realizing the full potential of this step-change in AI technology.

Now's the time for companies to use breakthrough advances in AI to set new performance frontiers—redefining themselves and the industries in which they operate.