# The AI in Music

# Why Large Language Model is important

- In 2022, the emergence of ChatGPT made knowledge acquisition much simpler.
- In 2023, Google's medical model surpassed professional doctors in image-based disease recognition.
- In 2024,
  - the arrival of Sora lowered the barrier for film production, enabling ordinary people to create blockbuster effects.
  - Suno made music creation possible without requiring professional knowledge.
  - AlphaFold predicted nearly all protein structures,
  - while Tesla's Optimus robot began performing simple labor tasks.
  - Devin AI engineers could autonomously write and test software.
- ……

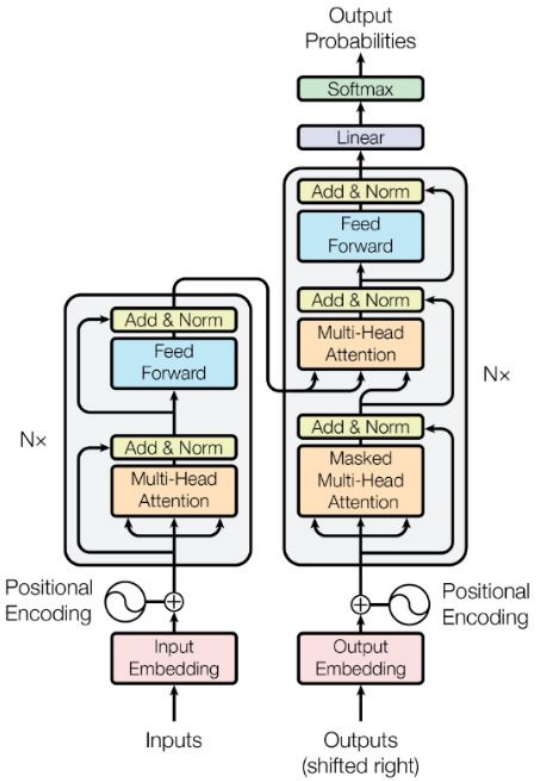
**Knowledge sharing will be replaced by large language models, robots will take over manual labor, and AI will handle basic software development. Mastering AI will be key to future opportunities.**

# Task for today

**Lecture 1**

- Learn the architecture of Transformers.
- Understand the development of large language models.
- Explore key models in the fields of music
- Deploy and try audiocraft

# The Transformers[1]




## The role of Transformers in large language models

### Revolutionary Breakthrough

The Transformer marked a major shift in NLP, moving from rule-based and statistical methods to deep learning. Its **self-attention mechanism** enables efficient and accurate handling of sequential data.

### Broad Applications

Transformers excel in NLP tasks like machine translation, text generation, and sentiment analysis, and have shown potential in fields like computer vision.

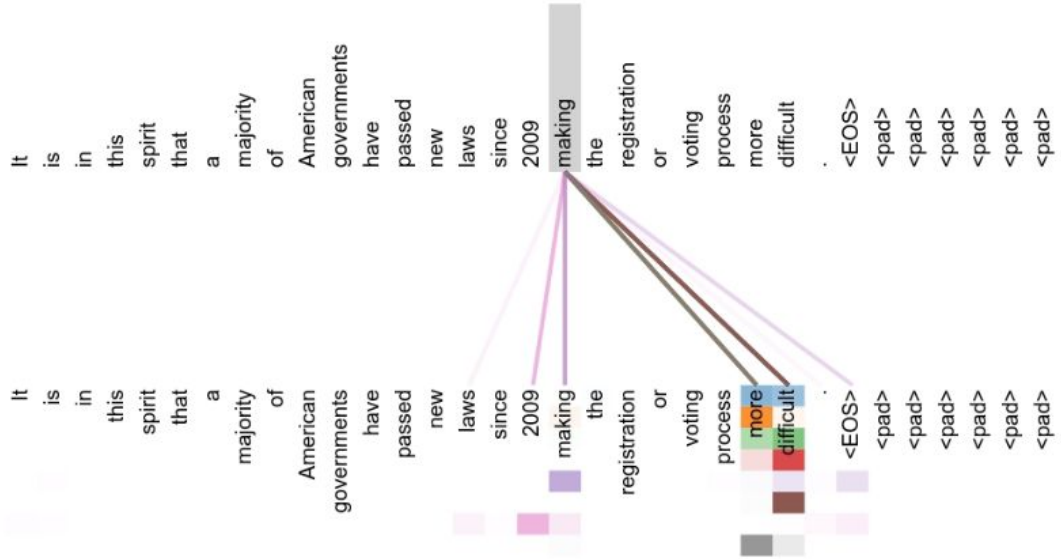
### Performance Advantages

Compared to RNNs and LSTMs, Transformers process data in parallel, improving efficiency and capturing long-distance dependencies critical for understanding complex language structures.

### Foundational Architecture

Transformers are the backbone of advanced models like GPT and BERT, driving both academic advancements and impactful commercial applications.

# Attention Visualization




**Dependencies in a Phrase**

- The word **"making"** attends strongly to distant words such as **"more"** and **"difficult"**.
- These words are crucial to complete the phrase **"making...more difficult"**.
- This illustrates how the Transformer captures the structure of long-distance dependencies in text, even when the related words are far apart.

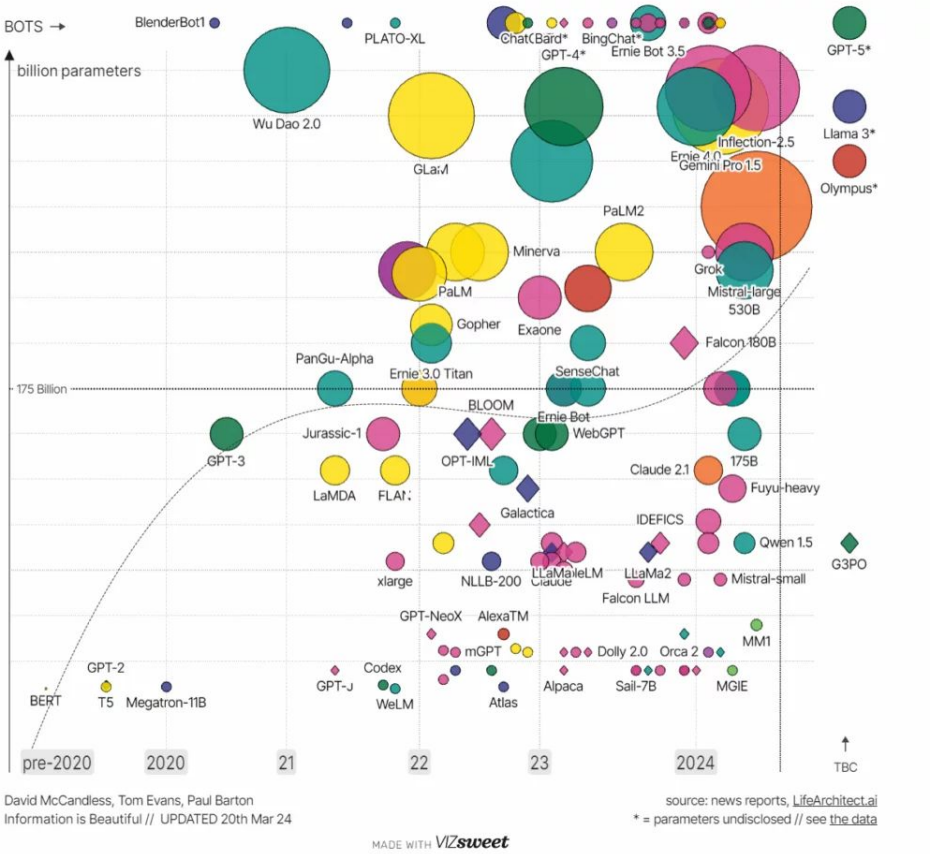
**Multi-Head Attention**

- Each color represents a different attention head.
- Different heads focus on different words or groups of words, allowing the model to capture multiple relationships simultaneously.
- For example:
  - One head focuses on "registration."
  - Another on "difficult."
  - Others might focus on contextual words like "process."

**Visual Representation**

- The height of the bars and the thickness of the lines represent the strength of attention (weights).
- Words with stronger attention contribute more to the contextual representation of "making."

# The Development of Large Language Models




**The rise and rise of AI-based Large Language Models (LLMs) like GPT4, LaMDA, LLaMa, PaLM and Jurassic-2.**

# Current Music Models(1)

- 1. **Suno AI**
  Generates realistic songs with vocals and instruments based on text prompts. Available since December 2023.
- 2. **Udio**
  Produces music from simple text prompts, including vocals and instrumentation. Free beta released in April 2024.
- 3. **TianGong Music**
  AI model optimized for Chinese music, supporting high-quality production and realistic vocal synthesis.
- 4. **Melodist**
  Open-source project focusing on AI-driven music composition, assisting in creating melodies and harmonies.
- 5. **AudioCraft by Facebook Research**
  Open-source tools for music generation, offering resources for training and fine-tuning models.

# Current Music Models(2)

- 6. **ChatMusician**
  Open-source text-to-music project enabling music generation from textual input.
- 7. **TianPuYue**
  Commercial AI music generation platform offering personalized song creation.
- 8. **QinYue by Tencent**
  AI model generating stereo audio or multi-track scores based on keywords or descriptive sentences.
- 9. **Seed-Music**
  Music generation and editing system supporting various creation tasks.
- 10. **Stable Audio by Stability AI**
  AI model producing high-quality tracks up to three minutes long, with audio-to-audio generation capabilities.
- 11. **HaiMian Music**
  AI platform allowing users to generate personalized music by inputting inspiration sentences, images, or existing lyrics.
- 12. **YinFeng**
  AI platform enabling complete song generation from lyrics, with options to add reference music or melody motifs.