



THE HONG KONG
POLYTECHNIC UNIVERSITY
香港理工大學



PolyU 理大商學院
Business School
Innovation-driven Education and Scholarship

School of
**ACCOUNTING
& FINANCE**
會計及金融學院

Week 11: Textual Analysis & Large Language Model in Accounting and Finance - Part 2

AF3214 Python Programming for Accounting and Finance

Vincent Y. Zhuang, Ph.D.
vincent.zhuang@polyu.edu.hk

School of Accounting and Finance
The Hong Kong Polytechnic University

R508, 8:30 am – 11:20 am, Wednesdays, Semester 2, AY 2024-25

Summary: how to train your GenAI - ChatGPT



every
~year

Stage 1: Pre-training

1. Download ~10TB of text.
2. Get a cluster of ~6,000 GPUs.
3. Compress the text into a neural network, pay ~\$2M, wait ~12 days.
4. Obtain **base model**.



<USER>

Can you write a short introduction about the relevance of the term "monopsony" in economics? Please use examples related to potential monopsonies in the labor market and cite relevant research.

<ASSISTANT>

"Monopsony" refers to a market structure where there is only one buyer for a particular good or service. In economics, this term is particularly relevant in the labor market, where a monopsony employer has significant power over the wages and working conditions of their employees. The presence of a monopsony can result in lower wages and reduced employment opportunities for workers, as the employer has little incentive to increase wages or provide better working conditions...

every
~week

Stage 2: Fine-tuning

1. Write labeling instructions
2. Hire people, collect 100K high quality ideal Q&A responses, and/or comparisons.
3. Finetune base model on this data, wait ~1 day.
4. Obtain **assistant model**.
5. Run a lot of evaluations.
6. Deploy.
7. Monitor, collect misbehaviors, fix, go to step 1.

Llama-2 models by Meta contain both base models and assistant models

LLM Leaderboard

LLM Leaderboard from “Chatbot Arena”, managed by UC Berkeley SkyLab
As of April, 2024

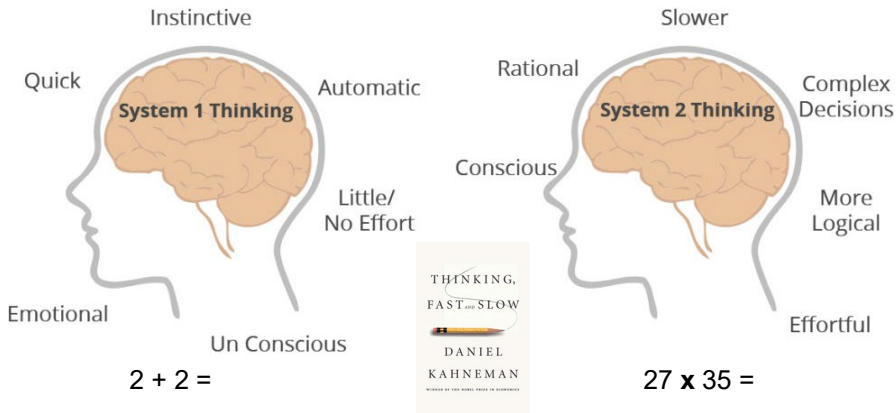
Rank ▲	Model ▲	★ Arena Elo ▲	📊 95% CI ▲	🗳 Votes ▲	Organization ▲	License ▲	Knowledge Cutoff ▲
1	GPT-4-Turbo-2024-04-09	1260	+5/-5	15751	OpenAI	Proprietary	2023/12
1	Claude_3_Opus	1255	+3/-4	56101	Anthropic	Proprietary	2023/8
1	GPT-4-1106-preview	1254	+3/-3	65159	OpenAI	Proprietary	2023/4
2	GPT-4-0125-preview	1250	+3/-4	50923	OpenAI	Proprietary	2023/12
5	Bard_(Gemini_Pro)	1209	+5/-5	12468	Google	Proprietary	Online
5	Claude_3_Sonnet	1203	+3/-3	62056	Anthropic	Proprietary	2023/8
7	Command_R+	1193	+4/-4	29437	Cohere	CC-BY-NC-4.0	2024/3
7	GPT-4-0314	1189	+4/-4	42925	OpenAI	Proprietary	2021/9
9	Claude_3_Haiku	1182	+3/-3	57727	Anthropic	Proprietary	2023/8
10	GPT-4-0613	1164	+3/-3	61520	OpenAI	Proprietary	2021/9
10	Mistral-Large-2402	1158	+3/-4	37650	Mistral	Proprietary	Unknown
11	Qwen1.5-72B-Chat	1154	+4/-5	27826	Alibaba	Qianwen LICENSE	2024/2
12	Claude-1	1150	+4/-5	21868	Anthropic	Proprietary	Unknown
12	Mistral_Medium	1148	+3/-5	30764	Mistral	Proprietary	Unknown
12	Command_R	1148	+3/-4	33061	Cohere	CC-BY-NC-4.0	2024/3

<https://chat.lmsys.org/?leaderboard>

Now about the future...

Types of thinking - 2 Systems

popularized by Thinking Fast and Slow

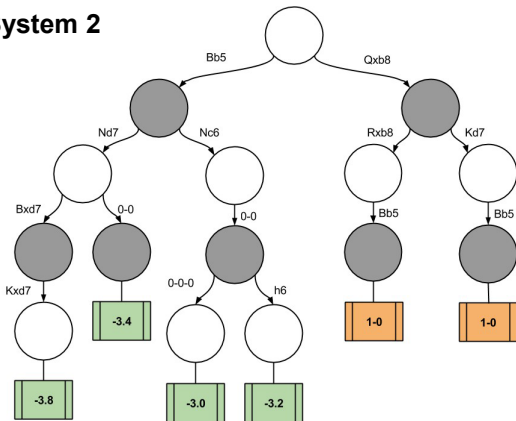


Let's switch gears to talk about some of the future directions of development in LLMs.

Now about the future...



System 2

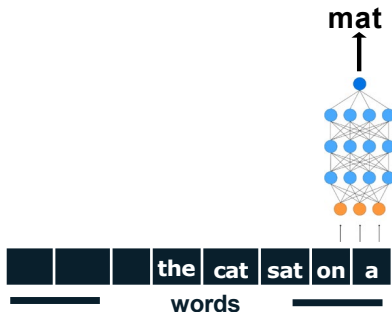


System 1: no time to think, instinctive move, generates the proposals (used in speed chess)

System 2: laying out trees of possibilities, keeping track of the tree, conscious effortful process (used in competitions)

Now about the future...

LLMs currently only have a System 1, very preliminary System 2 Reasoning

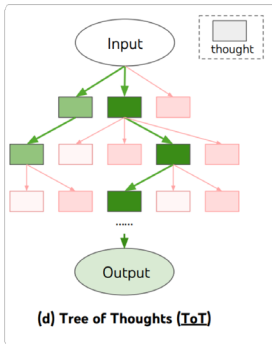


- Achieving human-level intelligence requires refining the transition from the fast & intuitive System 1 to slower System 2 reasoning.
- System 1 excels in quick, heuristic decisions, and System 2 relies on logical reasoning for more accurate judgments and reduced biases.
- Most LLMs excel at fast decision-making but lack the depth for complex reasoning.
- Recently, reasoning LLMs like OpenAI's o1/o3 and DeepSeek's R1 demonstrated expert-level performance in fields such as mathematics and coding, closely mimicking the deliberate reasoning of System 2 and showcasing human-like cognitive abilities.

Now about the future...

System 2

Like tree search in Chess, but in language.



We want to “think”: convert time (x-axis) to accuracy (y-axis).

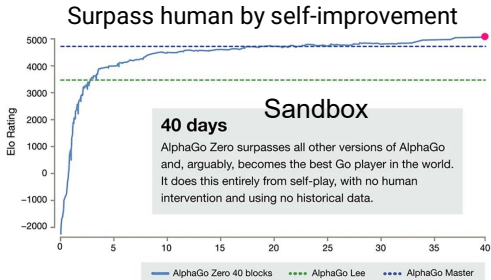
People are inspired by what to give LLMs a System 2.

Intuitively, we want to convert time into accuracy.

Now about the future...

Self-improvement

Go playing program by Deepmind



AlphaGo had two major stages (in 1st release):

1. Learn by imitating expert human players
2. Learn by self-improvement (reward = win the game)

Big question in LLMs:

What does Step 2 look like in the open domain of language? Main challenge: Lack of a reward criterion.

Now about the future...

Custom LLMs

GPTs app store:



[GPTs announcement at OpenAI dev day, Nov 2023]

Customize these LLMs and have them become experts at specific tasks;

Customization along the lines of specific custom instructions

Create a custom GPT:

< New GPT
+ Draft

Create Configure

Name
Name your GPT

Description
Add a short description about what this GPT does

Instructions
What does this GPT do? How does it behave? What should it avoid doing?

Conversation starters

Knowledge
Upload files

Capabilities
☒ Web Browsing
☒ DALL·E Image Generation
☐ Code Interpreter ⓘ

Actions
Add actions

Add knowledge by uploading files and ChatGPT does retrieval augmented generation

Reference chunks of that text in those files and use that when it creates responses

Now about the future...

LLM in a few years:

- It can read and generate text
- It has more knowledge than any single human about all subjects It can browse the internet
- It can use the existing software infrastructure (calculator, Python, mouse/keyboard)
- It can see and generate images and video It can hear and speak, and generate music
- It can think for a long time using a System 2
- It can “self-improve” in domains that offer a reward function
- It can be customized and fine-tuned for specific tasks, many versions exist in app stores It can communicate with other LLMs

The End