



Thinking in AI

— — — —
Gary Ang

4 questions and some hands-on

What is AI?

How to think about problems in AI?

What are some things to keep in mind when using Generative AI?

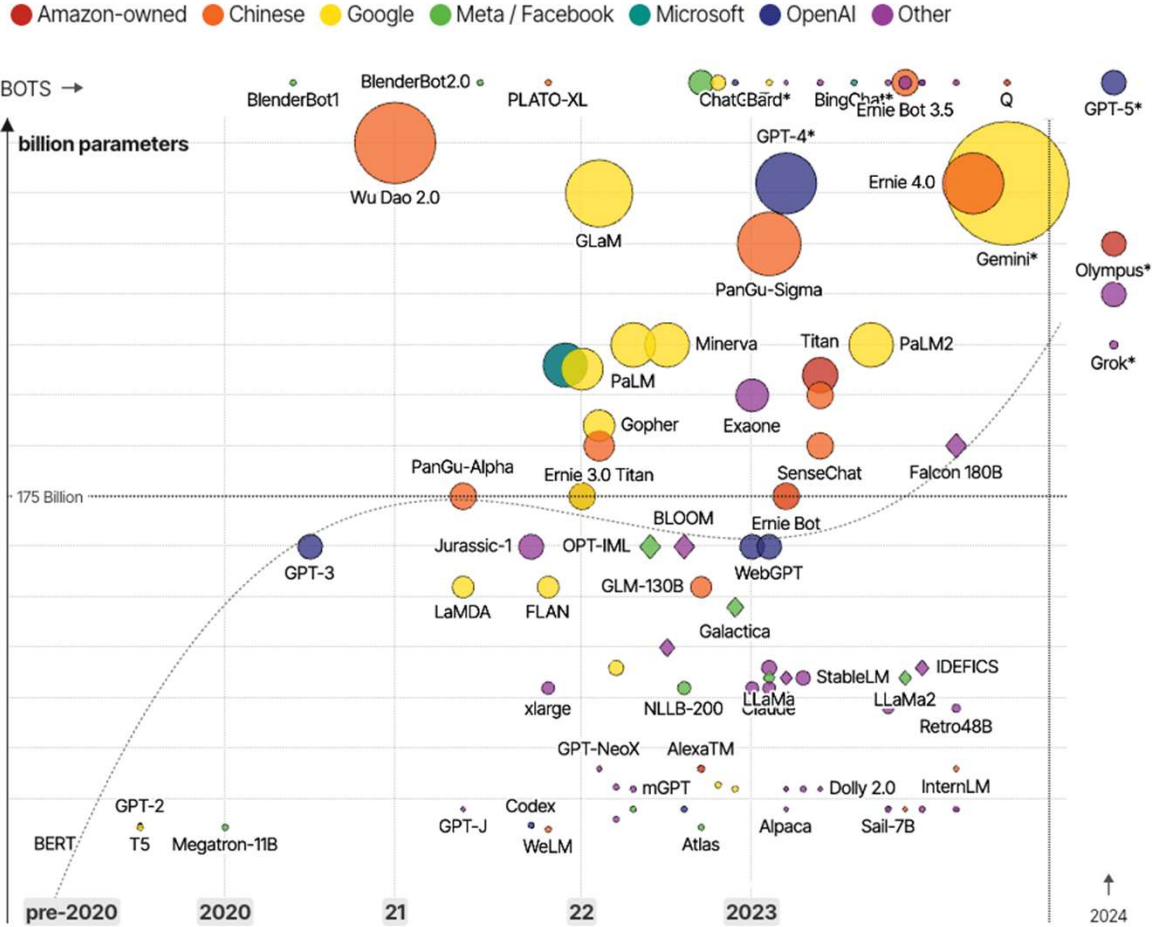
How can I keep up with AI developments?

Hands-on

AI and GenAI landscape changing every other day

New models trained & released; new tools developed ...

*Will not focus on the best and latest but ways of thinking to help
navigate rapidly evolving landscape*

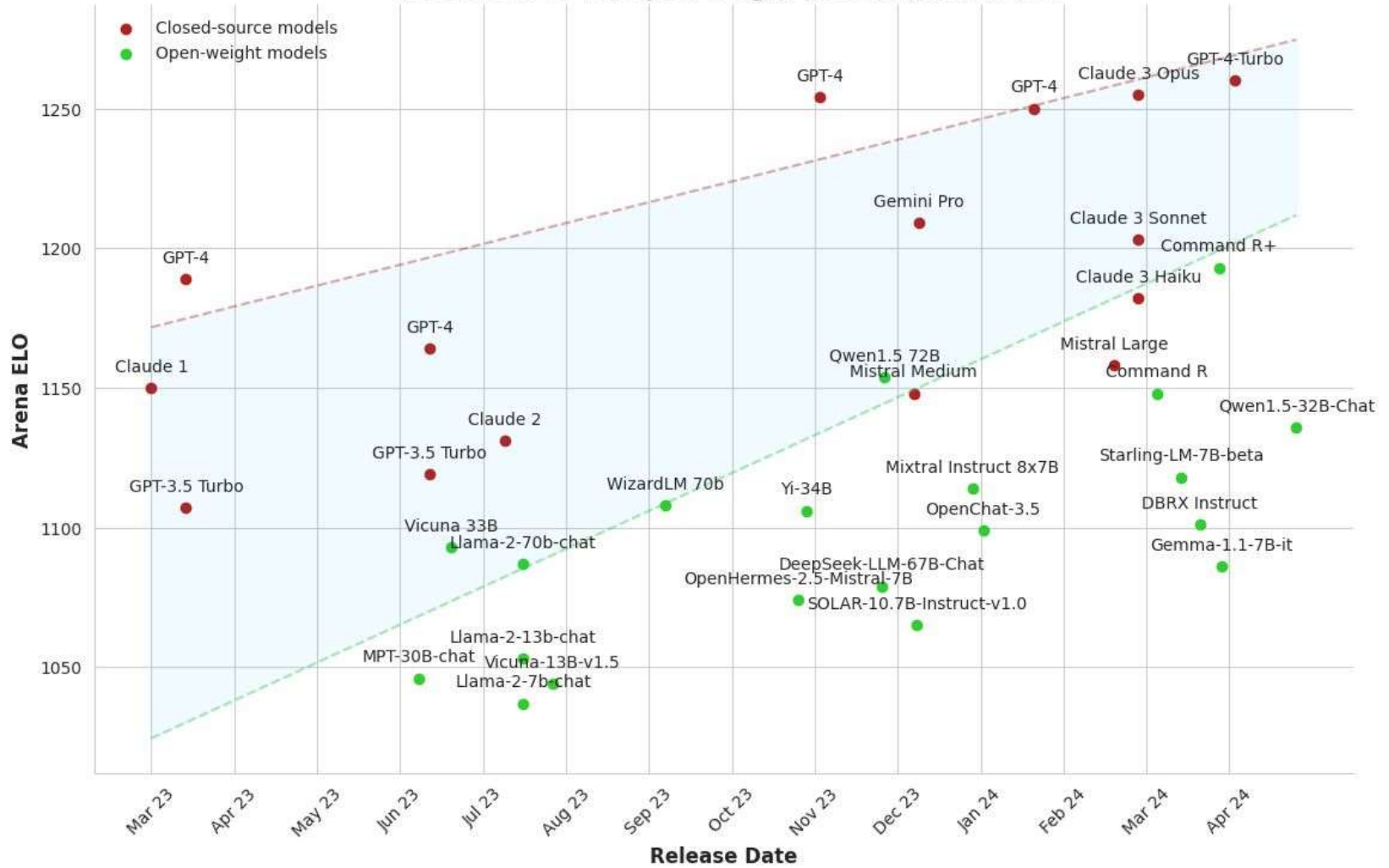


David McCandless, Tom Evans, Paul Barton
 Information is Beautiful // UPDATED 6th Dec 23

source: news reports, [LifeArchitect.ai](#)
 * = parameters undisclosed // see [the data](#)

Size of models over the years

Closed-source vs. Open-weight models (Arena ELO)



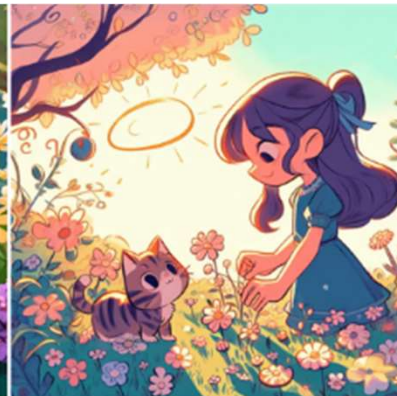
Progress of Closed Source vs Open Source LLMs, from Maxime Labonne's LinkedIn



MidJourney V5.2



Adobe Firefly 2



DALL-E 3

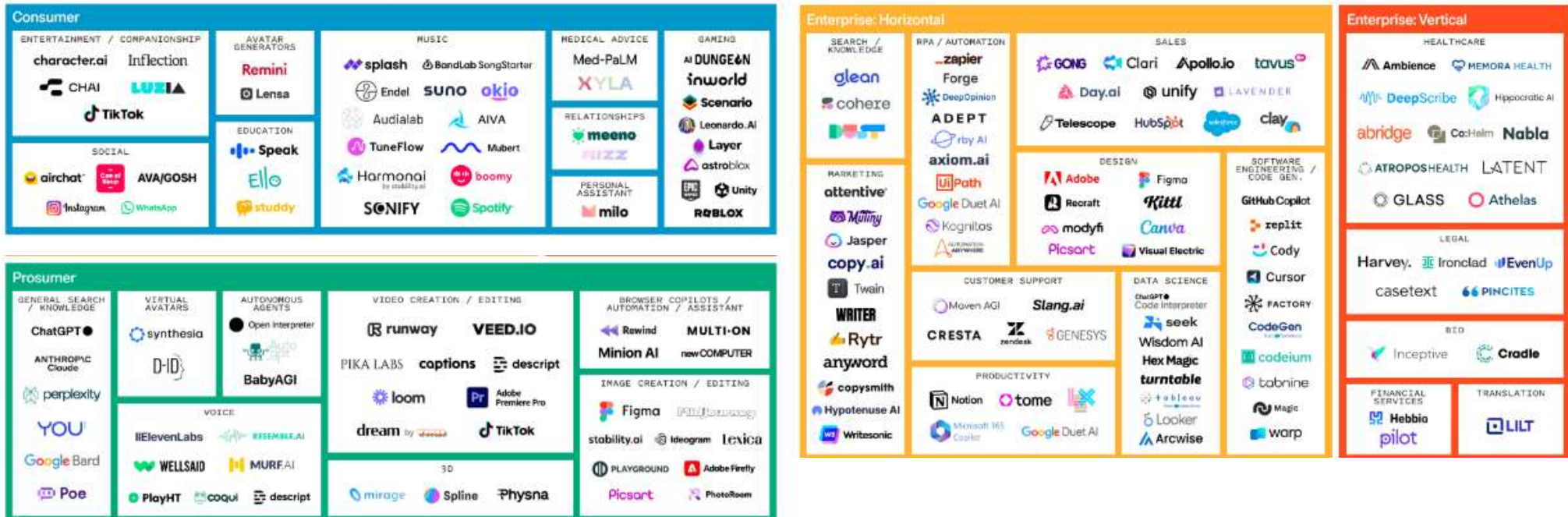


Stable Diffusion XL

Images generated on MidJourney V5.2, Adobe Firefly 2, DALL-E 3 and Stable Diffusion XL by Henrique Centieiro and Bee Lee

<https://levelup.gitconnected.com/midjourney-adobe-firefly-dall-e-stable-diffusion-which-ai-image-generator-should-you-choose-d40effe39c88>

Ecosystem of AI tools



Artificial Intelligence (AI)

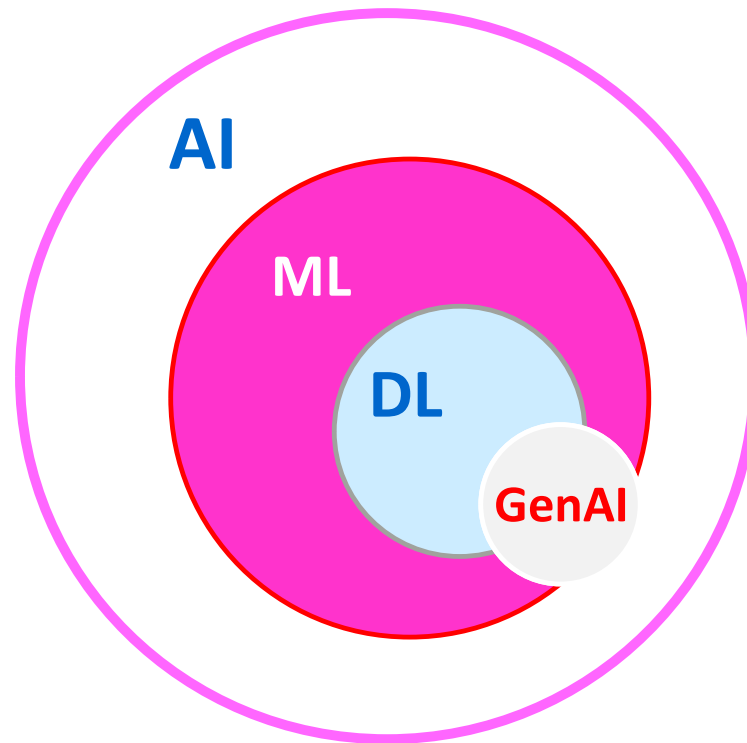
Machine Learning (ML)

Deep Learning (DL)

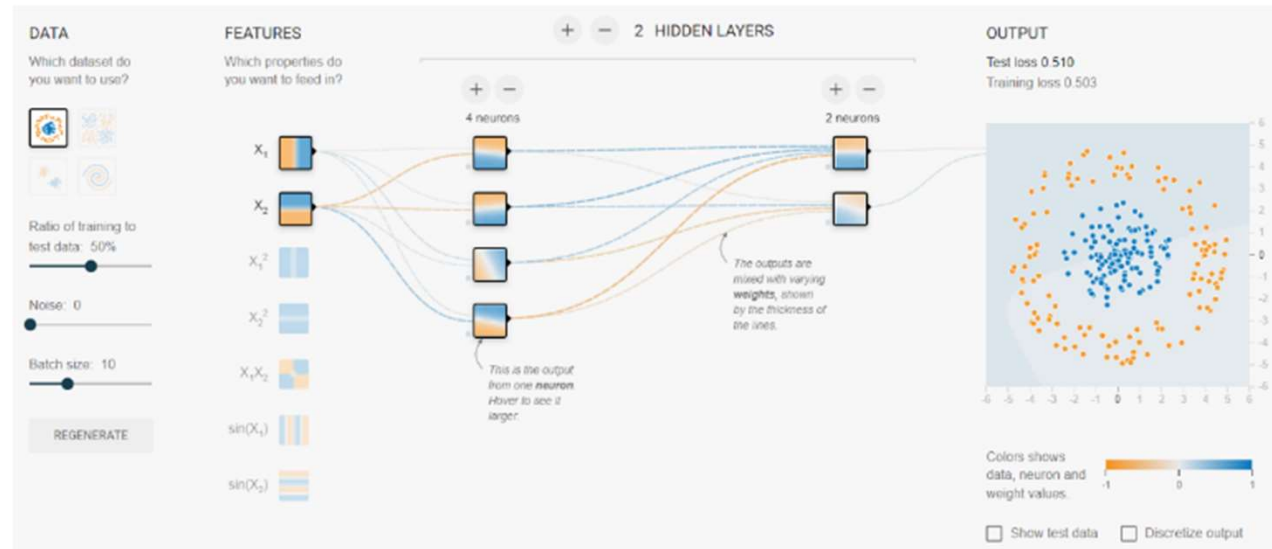
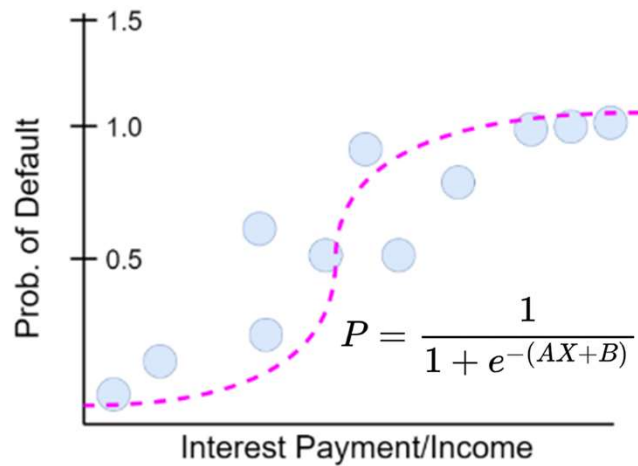
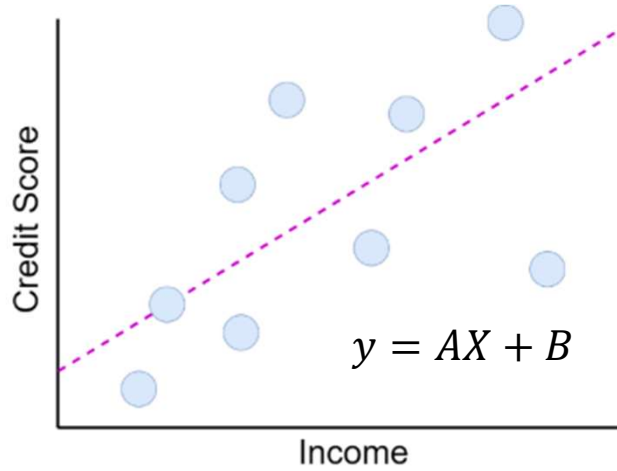
Generative AI Models (GenAI)

Language & image

Other modalities?

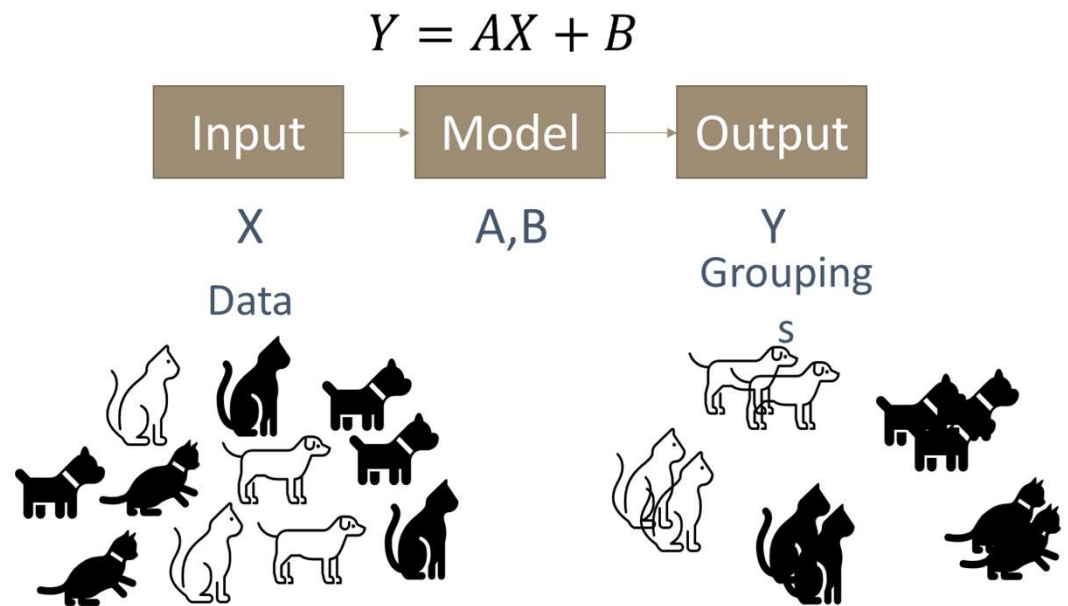
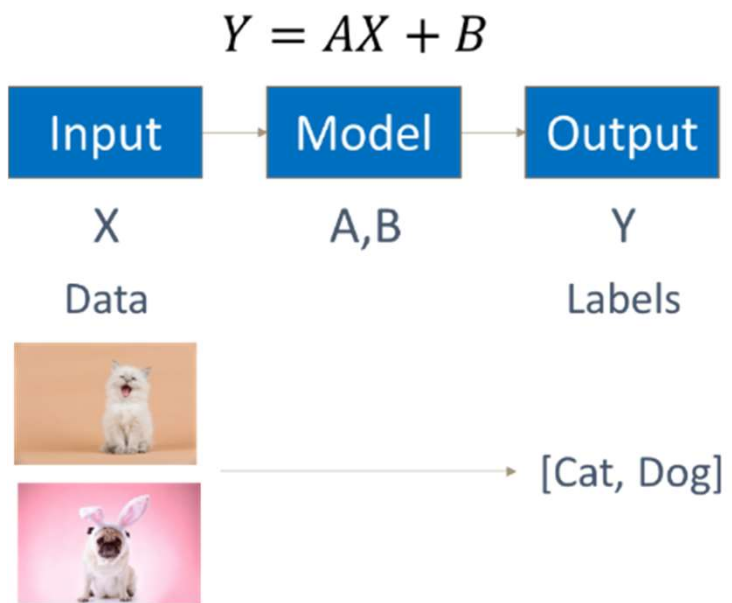


From Machine to Deep Learning



<https://playground.tensorflow.org/>

Supervised vs. unsupervised/self-supervised

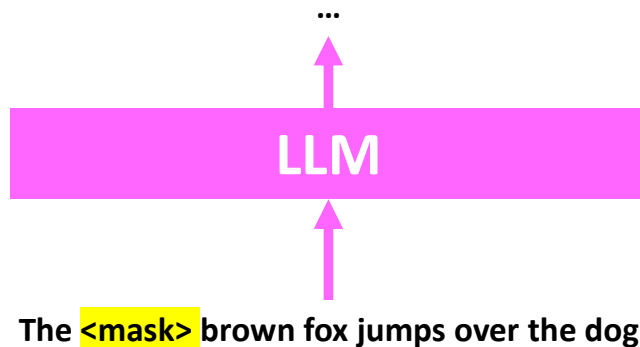


GenAI – Large Language Models (LLM) – Self Supervised

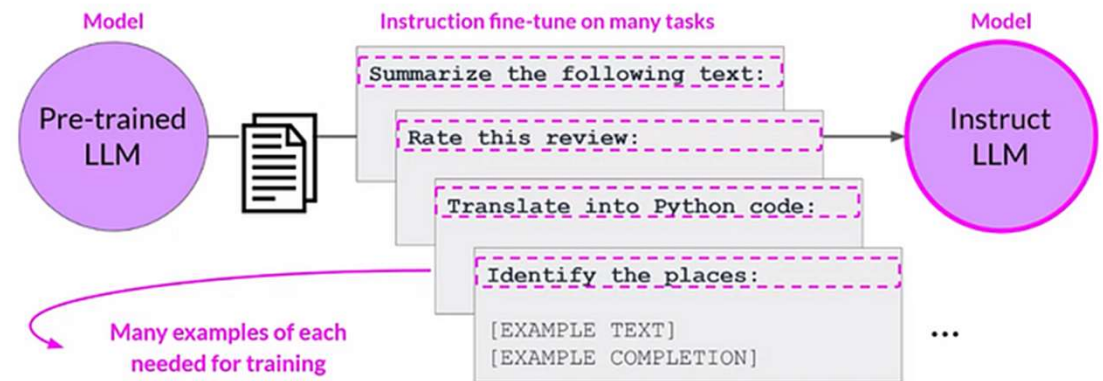
The **slow** brown fox jumps over the dog – 10%

The **fat** brown fox jumps over the dog – 5%

The **quick** brown fox jumps over the dog – 40%



Pretraining



Finetuning

Masking alone could already be viewed as different tasks

Social media sentiment – I missed the bus today. I felt so <Task output>

Translation - English: I love Paris. French: <Task output>

Question Answer – When did Singapore become independent?: <Task output>

Problem Framing for AI

- Decompose your problem into tasks – be **specific**
- Understand how well such tasks can be performed by AI/GenAI – be **realistic**
- Over time, develop a workflow and evaluate usefulness – be **systematic**
- Two useful (totally non-technical) resources, esp. when you move to more advanced usage of AI/GenAI
 - <https://developers.google.com/machine-learning/problem-framing>
 - <https://pair.withgoogle.com/guidebook>

Why prompt engineering?

- Unless you code or build your own models, prompts are main interface between humans and GenAI models.
 - Not an ideal interface, **extremely brittle!**
 - **Good to have some principles in mind**



Things to note

- **Prompts are model specific**
 - *Different models, different versions of models work with different prompts, e.g., GPT 4 vs. Claude Opus; gpt-4-turbo vs. gpt-4-turbo-2024-04-09, Dall-E 3 vs. Midjourney*
- **There is a limit to how much information you can stuff into prompts**
 - *4k tokens for GPT 3.5; 16k tokens for GPT 4; 128k tokens for GPT 4 turbo; 2m tokens for Gemini*
 - *1k tokens ~ 2 pages*
- **Not all information in prompts is equal**
 - *LLMs known to focus on start and end of prompts*
- **Hallucinations, hallucinations, hallucinations**
 - *DO NOT treat LLMs as experts, better to think of them as an intern or a powerful autocorrecting tool*

Basics

- **Creative, Balanced, Precise**

- *LLMs have settings like temperature, top-p that determine the degree of randomness*

- **Elements of a prompt**

- *Instruction, Context, Input Data, Output Indicator*

Classify the text into neutral, negative, or positive

<Examples>

I think the food was okay.

Sentiment:

Write me a social media post on CDC vouchers.

<Background information>

Post:

Approach

- **State the role** – SEO expert
- **State the task** - classify, summarise ...
- **Don't overload your query/request** – step by step, outline/plan then flesh out
- **Add structure** - <...>, ###...### (model and system specific)
- **Be specific and precise** – no need for pleasantries
- **Negations may not work** – don't generate a **hat** on the man
- **If don't know, say don't know** – reduces chances of hallucination

- *More examples*
 - <https://www.promptingguide.ai/introduction/examples>

LLMs for your domain

- LLMs may or may not have the knowledge or skills

- *Scope and nature of training data*
- *How it was trained*
- *Cut-off date*

Classify the sentiment.

This is awesome! // Negative

This is bad! // Positive

Wow that movie was rad! // Positive

What a horrible show! //

- **How to address?**

Not just limited to examples, what else?

Few Shot

Retrieval
Augmented
Generation

Finetuning

Techniques

- Other methods

- *Chain of thought*

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ❌

Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✅

Or just state “ Let’s think step by step.”

Many other methods but coding required

- *Graph of Thoughts*
- *Tree of Thoughts*
- *Chain of Knowledge*
- *Chain of Verification*
 - *Chain of Density*

Techniques

#Principle	Prompt Principle for Instructions
1	No need to be polite with LLM so there is no need to add phrases like “please”, “if you don’t mind”, “thank you”, “I would like to”, etc., and get straight to the point.
2	Integrate the intended audience in the prompt, e.g., the audience is an expert in the field.
3	Break down complex tasks into a sequence of simpler prompts in an interactive conversation.
4	Employ affirmative directives such as ‘do,’ while steering clear of negative language like ‘don’t’.
5	When you need clarity or a deeper understanding of a topic, idea, or any piece of information, utilize the following prompts: <ul style="list-style-type: none"> o Explain [insert specific topic] in simple terms. o Explain to me like I’m 11 years old. o Explain to me as if I’m a beginner in [field]. o Write the [essay/text/paragraph] using simple English like you’re explaining something to a 5-year-old.
6	Add “I’m going to tip \$xxx for a better solution!”
7	Implement example-driven prompting (Use few-shot prompting).
8	When formatting your prompt, start with ‘###Instruction###’, followed by either ‘###Example###’ or ‘###Question###’ if relevant. Subsequently, present your content. Use one or more line breaks to separate instructions, examples, questions, context, and input data.
9	Incorporate the following phrases: “Your task is” and “You MUST”.
10	Incorporate the following phrases: “You will be penalized”.
11	use the phrase “Answer a question given in a natural, human-like manner” in your prompts.
12	Use leading words like writing “think step by step”.
13	Add to your prompt the following phrase “Ensure that your answer is unbiased and does not rely on stereotypes”.
14	Allow the model to elicit precise details and requirements from you by asking you questions until he has enough information to provide the needed output (for example, “From now on, I would like you to ask me questions to...”).

Techniques

15	To inquire about a specific topic or idea or any information and you want to test your understanding, you can use the following phrase: "Teach me the [Any theorem/topic/rule name] and include a test at the end, but don't give me the answers and then tell me if I got the answer right when I respond".
16	Assign a role to the large language models.
17	Use Delimiters.
18	Repeat a specific word or phrase multiple times within a prompt.
19	Combine Chain-of-thought (CoT) with few-Shot prompts.
20	Use output primers, which involve concluding your prompt with the beginning of the desired output. Utilize output primers by ending your prompt with the start of the anticipated response.
21	To write an essay /text /paragraph /article or any type of text that should be detailed: "Write a detailed [essay/text /paragraph] for me on [topic] in detail by adding all the information necessary".
22	To correct/change specific text without changing its style: "Try to revise every paragraph sent by users. You should only improve the user's grammar and vocabulary and make sure it sounds natural. You should not change the writing style, such as making a formal paragraph casual".
23	When you have a complex coding prompt that may be in different files: "From now and on whenever you generate code that spans more than one file, generate a [programming language] script that can be run to automatically create the specified files or make changes to existing files to insert the generated code. [your question]".
24	When you want to initiate or continue a text using specific words, phrases, or sentences, utilize the following prompt: o I'm providing you with the beginning [song lyrics/story/paragraph/essay...]: [Insert lyrics/words/sentence]'. Finish it based on the words provided. Keep the flow consistent.
25	Clearly state the requirements that the model must follow in order to produce content, in the form of the keywords, regulations, hint, or instructions
26	To write any text, such as an essay or paragraph, that is intended to be similar to a provided sample, include the following instructions: o Please use the same language based on the provided paragraph[/title/text /essay/answer].

PROMPT ENGINEERING FUNDAMENTALS

LARGE LANGUAGE MODELS
 GENERATIVE AI APPLICATIONS
 FEW SHOT LEARNING
 CHAT COMPLETION

CAN AI HELP SCHOOL ADMIN WITH SCHEDULES?
 CAN TEACHERS USE IT TO PLAN LESSONS?
 CAN IT BE MY TUTOR?

EDUCATION ENTREPRENEURS

1 CORE CONCEPTS

DAY 01

TOKENIZED PROMPT

HOW DO PROMPTS WORK IN LLM?

COMPLETION PREDICTED NEXT TOKEN

SLIDING CONTEXT WINDOW

PROMPT

NEXT COMPLETION

- ✓ PROMPT & TOKENIZATION
- ✓ BASE LLM BEHAVIOR
- ✓ INSTRUCTION-TUNED LLM
- ✓ CHAT COMPLETION EXAMPLE

2 CORE CHALLENGES

DAY 02

PROMPT

LLM

LLM CAN GENERATE DIVERGENT RESPONSES FOR SAME PROMPT

STOCHASTIC MODEL RESPONSE

MODEL RESPONSE FABRICATION

MODEL CAPABILITY VARIATION

MODEL TOKEN LIMITS & COST

LLM CAN BE USED TO PREDICT AN OUTCOME FROM TEXT

INSTRUCTION

WE CAN PROMPT AN LLM TO FOLLOW A SPECIFIC INSTRUCTION

APPLICABLE TO MANY USE CASES

DIFFERENT COSTS FOR DIFFERENT MODELS

NEW WORKS TO BE DONE

- ✓ STOCHASTIC MODEL RESPONSE
- ✓ MODEL RESPONSE FABRICATION
- ✓ MODEL CAPABILITY VARIATION
- ✓ MODEL TOKEN LIMITS & COST

3 GITHUB COPILOT CASE STUDY

DAY 03

CODE COMPLETION, CODE REVISIONS, CODE CHANGES

VISUAL STUDIO CODE EXTENSION

- ✓ BASE MODEL : TEXT GENERATION
- ✓ FINE-TUNED : CODE GENERATION
- ✓ GITHUB COPILOT : LLM APPLICATION
- ✓ INSIGHTS FOR : PROMPT DESIGN

4 PROMPT CONSTRUCTION

DAY 04

PROMPT

PROMPT SET

A NOTEBOOK

BASIC PROMPTS ALLOW MODEL TO COMPLETE AS SINGLY ACTION

INSTRUCTION PROMPTS ADD SPECIFIC CONTEXT OR CONTENT TO INFLUENCE RESULT

PERSONA

YOU ARE A FISH ASSISTANT WHO ANSWERS TO QUESTIONS

- ✓ BASIC PROMPT : COMPLETION
- ✓ COMPLEX PROMPT : MULTI-TURN
- ✓ INSTRUCTION PROMPT : SET TASK
- ✓ SYSTEM CONTEXT : SET PERSONA

5 PRIMARY CONTENT

DAY 05

FEW SHOT LEARNING

GIVE EXAMPLES OF RESPONSES - LLM DETECTS A PATTERN FOR YOU

POUNDS → CRICKET

OUTS → BASKETBALL

BOUNDBOARDS → FOOTBALL

DRINKS →

→ BASKETBALL

TEMPLATE FOR RECIPE

LEARN FROM THIS!

- ✓ DESIGN PATTERN FOR ACTION
- ✓ AS EXAMPLES → FEW SHOT LEARNING
- ✓ AS CUES → PRIME THE RESPONSE
- ✓ AS TEMPLATES → REUSABLE 'RECIPES'

6 ADVANCED TECHNIQUES

DAY 06

THINK ABOUT DESIGN PATTERNS FOR PROMPTS

CHAIN-OF-THOUGHT (PATTERN)

SELF-REFINE (CRITIQUING)

GENERATED KNOWLEDGE (YOUR DATA)

VARY PARAMETERS (TEMPERATURE)

AND EXPERIMENT WITH TUNING THE LLM PARAMETERS

TEMPERATURE OF MODEL AFFECTS THE RESPONSE CREATIVITY

7 BEST PRACTICES

DAY 07

REFINE PROMPT FOR YOUR DATA & APP DOMAIN

ITERATE PROMPTS & EVALUATE QUALITY BY TRYING DIVERSE TECHNIQUES - BUILD A WORKFLOW

USE THE GENERATED MODEL TO MONITOR PERFORM

IDEATION

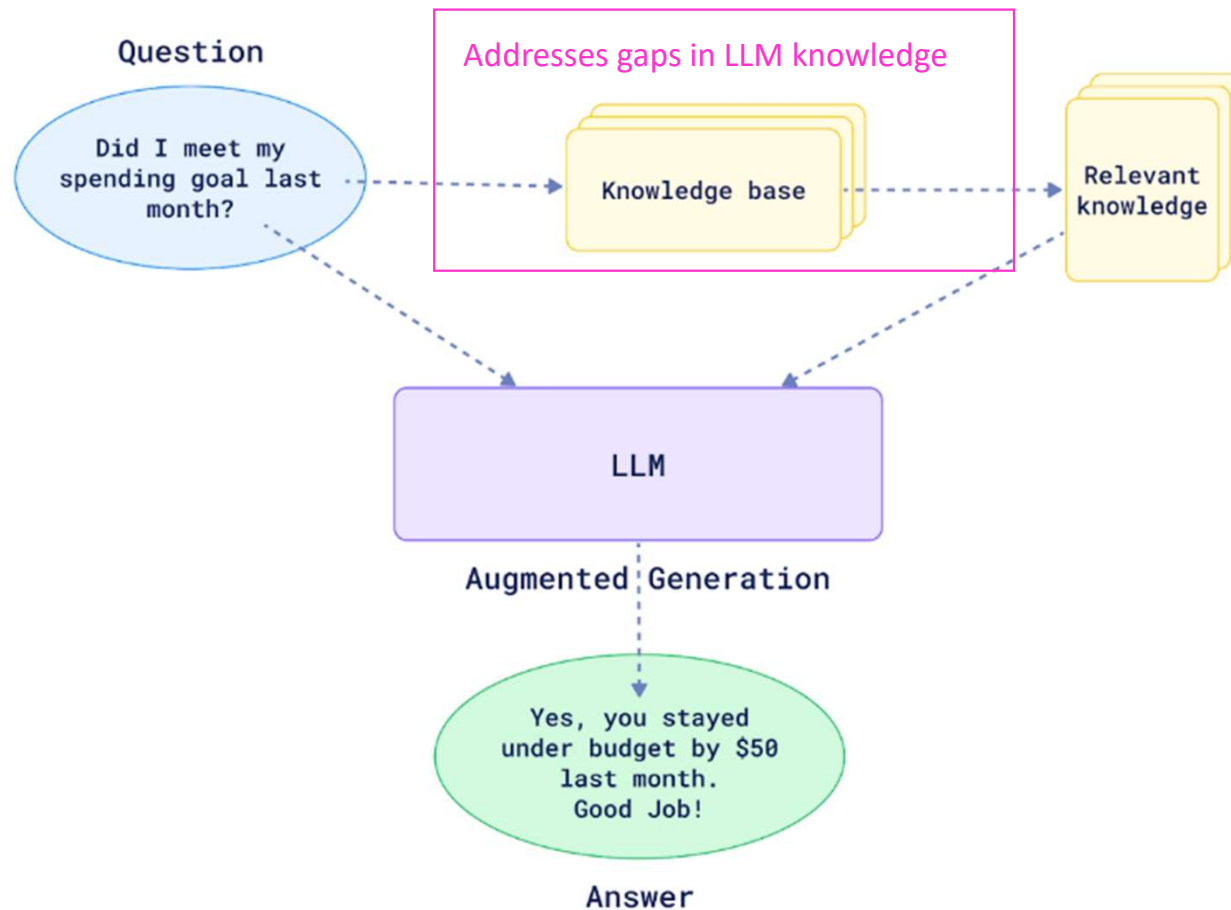
AUGMENTATION

OPERATIONALIZATION

- ✓ DOMAIN UNDERSTANDING
- ✓ MODEL UNDERSTANDING
- ✓ ITERATION & VALIDATION
- ✓ STREAMLINED WORKFLOW

SKETCH NOTES BY NITYA

Retrieval Augmented Generation (RAG)



Retrieval Augmented Generation (RAG)

My GPTs

+ Create

GPTs

Discover and create custom versions of ChatGPT that combine instructions, extra knowledge, and any combination of skills.

🔍 Search GPTs

Top Picks Writing Productivity Research & Analysis Education Lifestyle Programming

Featured

Curated top picks from this week



Landing Page Creator from HubSpot

Generate landing pages for your next marketing campaign. Edit and publish your page in minutes with...
By hubspot.com



Zumper Rentals - Apartments and...

Your friendly US & Canada rental home search assistant. Let me help you pick the perfect neighborhood...
By zumper.com

Retrieval Augmented Generation (RAG)

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

If you upload files under Knowledge, conversations with your GPT may include file contents. Files can be downloaded when Code Interpreter is enabled

Upload files

Capabilities

- Web Browsing
- DALL-E Image Generation
- Code Interpreter & Data Analysis ⓘ

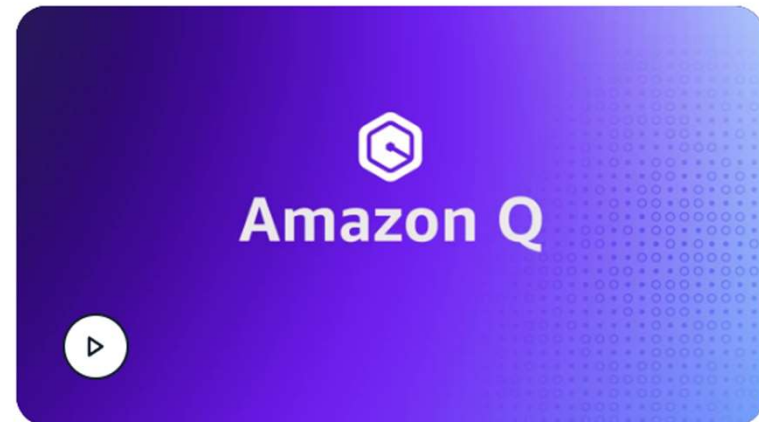
Amazon Q (Preview)

Your generative AI-powered assistant designed for work that can be tailored to your business

Available plans

Welcome to a new world of work with Amazon Q

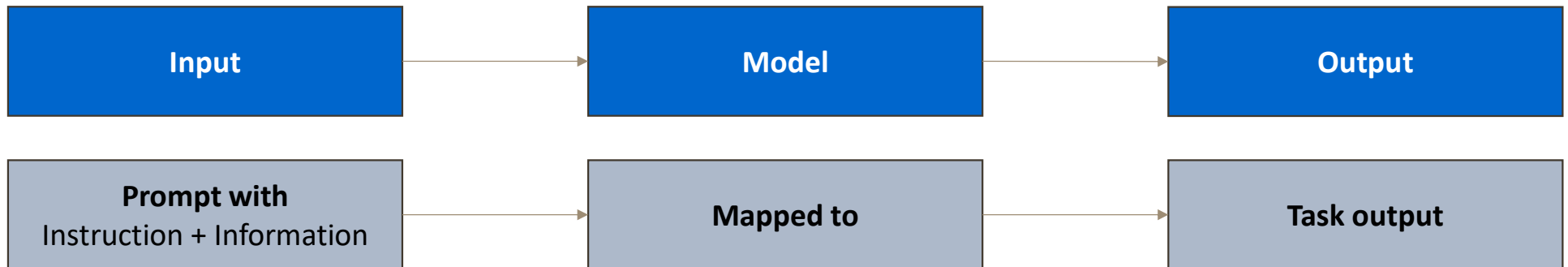
Amazon Q can help you get fast, relevant answers to pressing questions, solve problems, generate content, and take actions using the data and expertise found in your company's information repositories, code, and enterprise systems. When you chat with Amazon Q, it provides immediate, relevant information and advice



A large, fluffy white cloud is centered in the upper half of the image. A silver ladder extends from the bottom center of the frame up to the base of the cloud. The background is a solid, clear blue sky. The text "Things to take note of" is superimposed on the cloud in a bold, black, sans-serif font.

Things to take note of

GenAI ≠ Human – Try to not anthropomorphize GenAI, it's just math (for now)

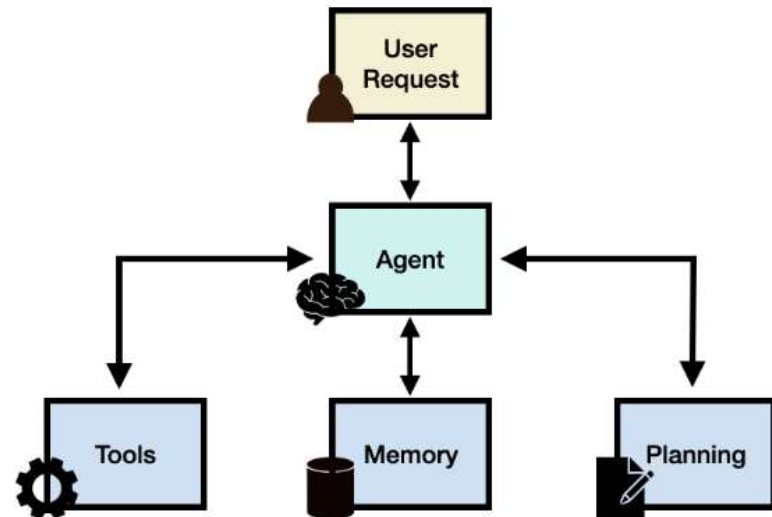
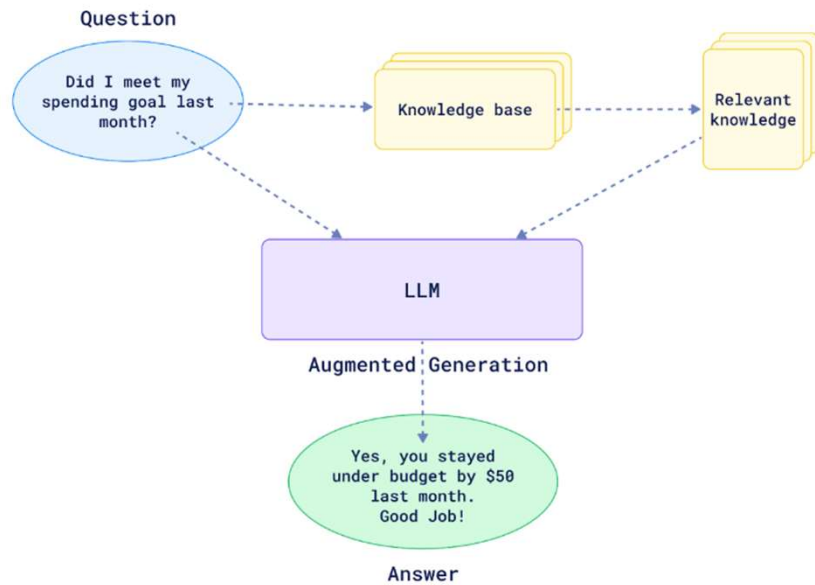


- Is the domain known to the model?
 - Is this something the model understands?
 - Text vs. vision vs. multimodal models
 - Have things changed since model was trained?
 - How much information can the model capture or remember?
- Are these tasks that the model may have been incidentally or deliberately trained to do?
 - Is task known to be easy or difficult for the model?


GenAI Model \neq GenAI System

Most GenAI/LLM chatbots now have some of these features under the hood, but these are design choices.

ChatGPT, Copilot, POE etc. may implement these differently.



Negations

 You

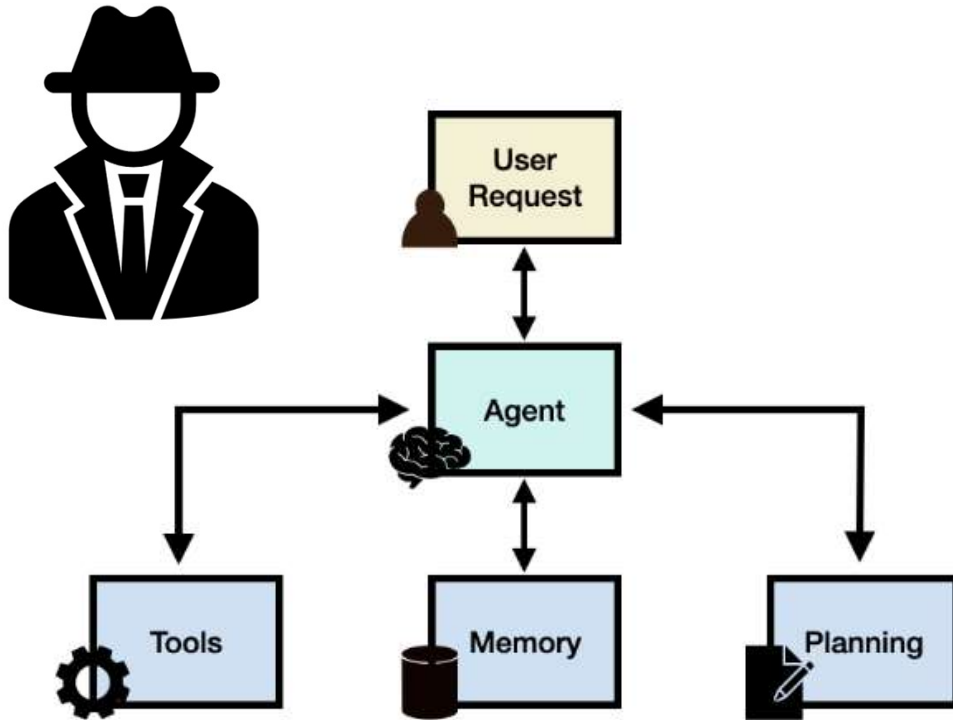
Chinese man not wearing glasses working from home.

 Designer

I've created an image of a Chinese man working from home without wearing glasses. I hope it captures the essence of a cozy and productive home workspace!



Agents

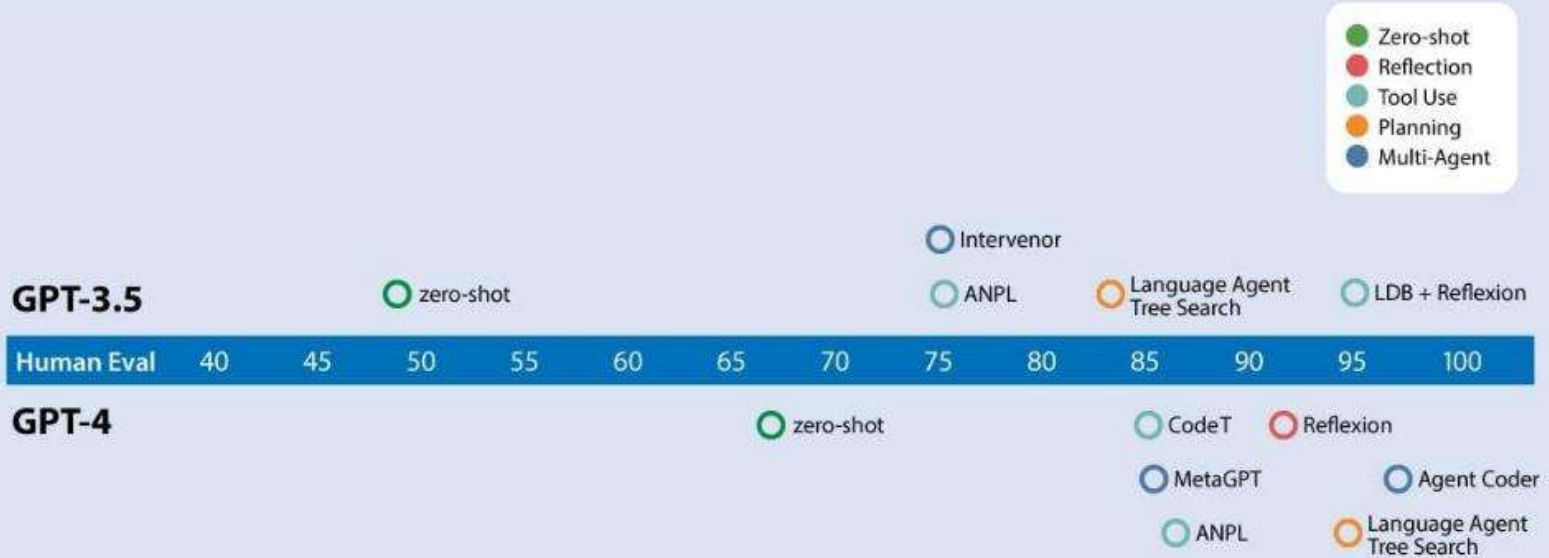


Example

- *Plan an outline.*
- *Decide what, if any, web searches are needed to gather more information.*
- *Write a first draft.*
- *Read over the first draft to spot unjustified arguments or extraneous information.*
- *Revise the draft taking into account any weaknesses spotted.*
- *And so on.*

Agents

GPT-3.5 and GPT-4 performance using zero-shot and agent workflows



Performance of GPT-3.5 and GPT-4 (zero-shot) on HumanEval, along with algorithms that use agent workflows on top of GPT-3.5 or GPT-4. Thanks to Joaquin Dominguez and John Santerre for help with this analysis.

Even if you cannot code, can act like an agent when you use LLMs

Agentic Design Patterns: Reflection



Please write code for {task}

```
def do_task(x): ...
```

```
def do_task_v2(x):
```

```
def do_task_v3(x):
```



Coder Agent
(LLM)

There's a bug on line 5. Fix it by ...

It failed Unit Test 3. Try changing ...



Critic Agent
(LLM)

Even if you cannot code, can act like an agent when you use LLMs

Agentic Design Patterns: Planning



image.jpg



final.jpg

Pose Determination

openpose model

Pose-to-Image

google/vit model

Example adapted from "HuggingGPT: Solving AI Tasks with ChatGPT and its Friends in Hugging Face," Shen et al. (2023)

Even if you cannot code, can act like an agent when you use LLMs

Agentic Design Patterns: Tool Use

Web search tool	Code execution tool
<p>You</p> <p>What is the best coffee maker according to reviewers?</p> <p>Copilot</p> <p>Searching for best coffee maker according to reviewers</p>	<p>You</p> <p>If I invest \$100 at compound 7% interest for 12 years, what do I have at the end?</p> <pre>principal = 100 interest_rate = 0.07 years = 12 value = principal*(1 + interest_rate)**years</pre>
Example from Bing CoPilot	Example from ChatGPT

Use social media to keep up




Chase Lean @chaseleanj

AI educator. I share practical
Sponsorships →
331 Following 60.5K Followers
Followed by ArtMin

Posts Replier

Chase Lean @chaseleanj · Midjourney version
I compared it to the categories.
Here are the result:
198

Chase Lean reposted
Chase Lean @chaseleanj · I've been experimenting with AI.



elvis @omarsar0


Building with LLMs @dair.ai · Prev: Meta AI, Galactica LLM, PapersWithElastic, PhD · Creator of the Prompting Guide (3.5M+ learners)
Science & Technology linktr.ee/elvissaravia · Joined Sept
477 Following 187.4K Followers
Followed by Muratcan Koylan, Simo Ryu, and 51 others you follow

Posts Replies Highlights Media

elvis @omarsar0 · Feb 1
Prompt Engineering Guide reaches 40K!

Surreal moment to be honest. I've been so passionate about democratizing AI research and education since I started out in this space.

Working tirelessly hours together with the open-source community and @dair.ai is the work I am most proud...
[Show more](#)




Yohei @yoheinakajima

VC by day, builder by night: @public log: yohei.me
Venture Capital Seattle
Joined April 2009
8,378 Following 70.2K Followers
Followed by Muratcan Koylan

Posts Replies

Yohei @yoheinakajima ·
...



Jerry Liu @jerryliu0

co-founder/CEO @llama_index

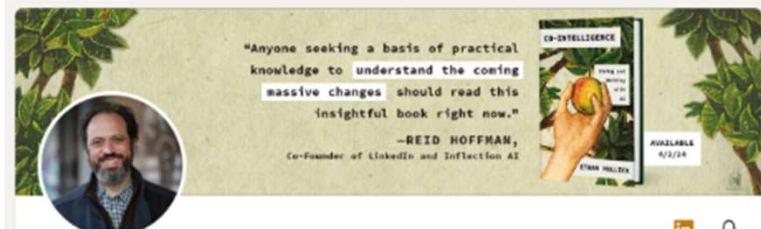
Careers: llamaindex.ai/careers
Enterprise: llamaindex.ai/contact
@llamaindex.ai · Joined September 2011
1,329 Following 43.2K Followers
Followed by Omar Khattab, LlamaIndex, and 44 others you follow

Posts Replies Highlights Media Likes

Jerry Liu @jerryliu0 · Mar 26
We're looking for an applied AI engineer! 🙌

If you're an AI engineer and you've specifically built deployed LLM/RAG/agent applications to customers and really like doing so, come talk to us at @llama_index.


Building production-quality LLM/RAG applications is challenging. We're...
[Show more](#)



Ethan Mollick · 3rd

Associate Professor at The Wharton School. Author of Co-Intelligence
Top Voice
United States · Contact info
Pre-order Co-Intelligence
105,418 followers
Followed by Marcus Lim, Tara Tan, and 4 others

The Wharton School
MIT Sloan School of Management



Eduardo Ordax · 2nd

Generative AI Lead @ AWS | Startup Advisor | Public Speaker
Madrid, Community of Madrid, Spain · Contact info
aws.amazon.com
22,187 followers · 500+ connections
Message Following More

Highlights

MLOps - E11 - Generative AI in Enterprise
Eduardo spoke at this event

About

One of my favorite quotes that I strive to live by every day is, "To give anything less than your best is to sacrifice the gift." In addition to this guiding principle, I have a daily ritual that I cherish - rising early in the morning to run a few miles, setting the tone for an energized and productive day ahead."
...see more

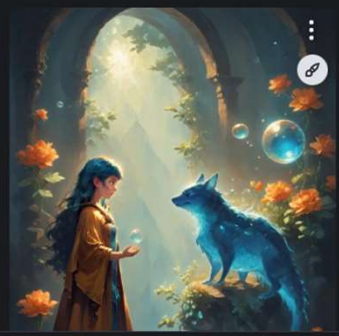
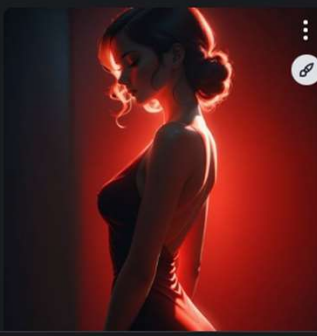
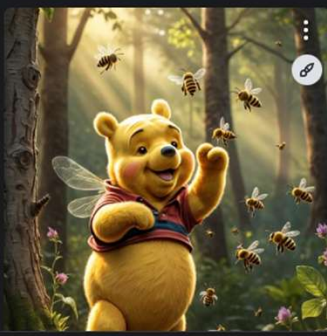
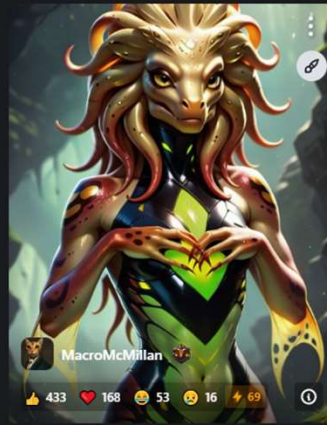
Resources to Start With

- <https://github.com/aishwaryanr/awesome-generative-ai-guide/>
- <https://github.com/microsoft/generative-ai-for-beginners>
- <https://www.promptingguide.ai/>
- <https://www.deeplearning.ai/>
- <https://lilianweng.github.io/posts/2023-03-15-prompt-engineering/>
- <https://cookbook.openai.com/>
- <https://arxiv.org/abs/2312.16171v1>
- <https://twitter.com/chaseleantj>
- <https://civitai.com/>
- <https://huggingface.co/>

Featured Images

All sorts of cool pictures created by our community, from simple shapes to detailed landscapes or human faces. A virtual canvas where you can unleash your creativity or get inspired.

Explore all images →



Hugging Face Search models, datasets, users... Models Datasets Spaces Posts Docs Solutions Pricing

Tasks Libraries Datasets Languages Licenses Other

Filter Tasks by name

Multimodal

- Image-Text-to-Text
- Visual Question Answering
- Document Question Answering

Computer Vision

- Depth Estimation
- Image Classification
- Object Detection
- Image Segmentation
- Text-to-Image
- Image-to-Text
- Image-to-Image
- Image-to-Video
- Unconditional Image Generation
- Video Classification
- Text-to-Video
- Zero-Shot Image Classification
- Mask Generation
- Zero-Shot Object Detection
- Text-to-3D
- Image-to-3D
- Image Feature Extraction

Natural Language Processing

- Text Classification
- Token Classification
- Table Question Answering
- Question Answering
- Zero-Shot Classification
- Translation
- Summarization
- Feature Extraction
- Text Generation
- Text2Text Generation
- Fill-Mask
- Sentence Similarity

Audio

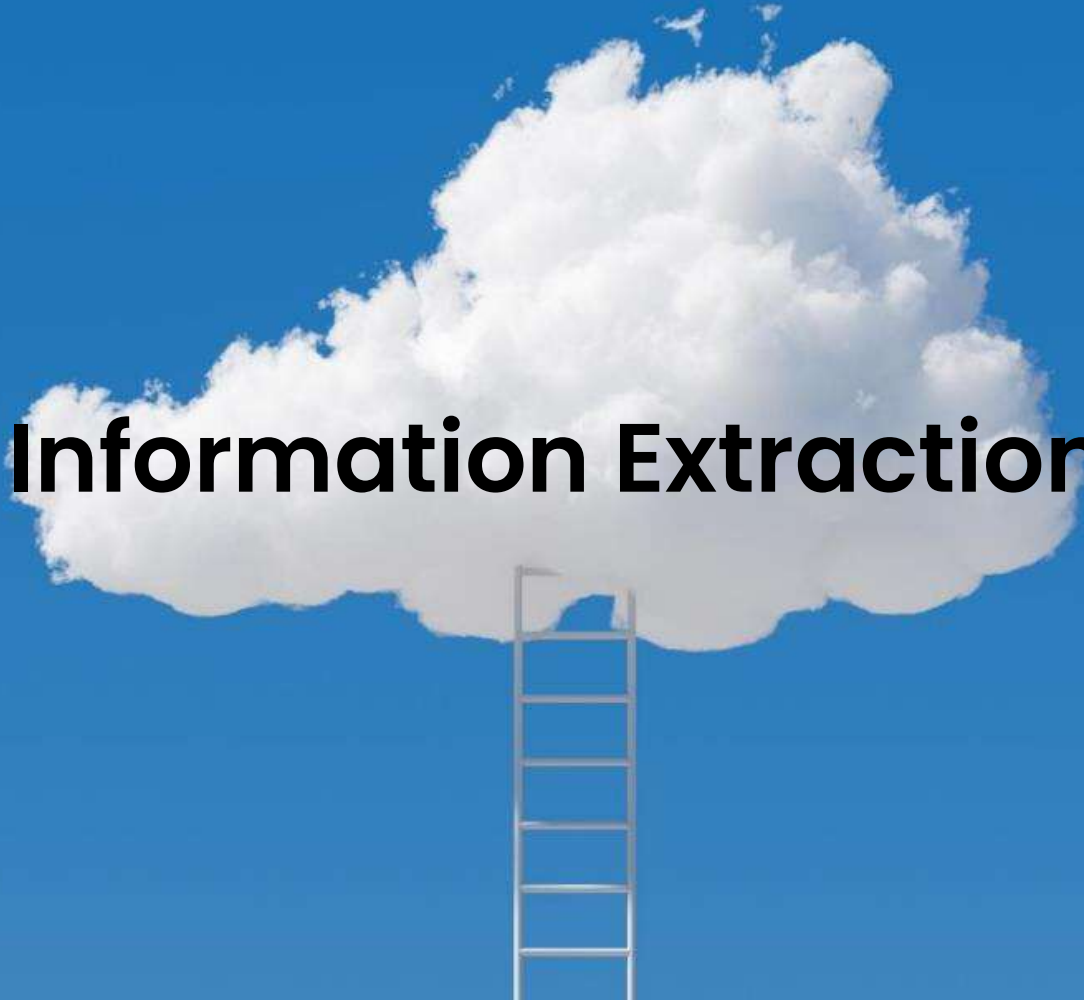
- Text-to-Speech
- Text-to-Audio
- Automatic Speech Recognition
- Audio-to-Audio
- Audio Classification
- Voice Activity Detection

Models 608,256 Filter by name new Full-text search Sort: Most likes

- runwayml/stable-diffusion-v1-5**
Text-to-Image • Updated Aug 24, 2023 • 4.07M • 10.7k
- CompVis/stable-diffusion-v1-4**
Text-to-Image • Updated Aug 24, 2023 • 1.28M • 6.27k
- stabilityai/stable-diffusion-xl-base-1.0**
Text-to-Image • Updated Oct 31, 2023 • 4.58M • 4.96k
- bigscience/bloom**
Text Generation • Updated Jul 29, 2023 • 8.24k • 4.54k
- meta-llama/Llama-2-7b**
Text Generation • Updated about 4 hours ago • 3.83k
- stabilityai/stable-diffusion-2-1**
Text-to-Image • Updated Jul 6, 2023 • 1.73M • 3.69k
- mistralai/Mixtral-8x7B-Instruct-v0.1**
Text Generation • Updated Feb 29 • 822k • 3.66k
- WarriorMama777/OrangeMixs**
Text-to-Image • Updated Jan 7 • 5.07k • 3.66k
- lillyasviel/ControlNet**
Updated Feb 25, 2023 • 3.46k
- meta-llama/Llama-2-7b-chat-hf**
Text Generation • Updated about 4 hours ago • 1.47M • 3.37k
- lillyasviel/ControlNet-v1-1**
Updated Apr 26, 2023 • 3.24k
- mistralai/Mistral-7B-v0.1**
Text Generation • Updated Dec 12, 2023 • 2.05M • 3.1k
- microsoft/phi-2**
Text Generation • Updated Feb 6 • 864k • 3.07k
- prompthero/openjourney**
Text-to-Image • Updated May 16, 2023 • 68.5k • 3.04k
- google/gemma-7b**
Text Generation • Updated about 19 hours ago • 174k • 2.79k
- THUDM/chatglm-6b**
Updated Sep 4, 2023 • 26.3k • 2.77k
- bigcode/starcoder**
Text Generation • Updated 27 days ago • 14.3k • 2.7k
- CompVis/stable-diffusion-v1-4-original**
Text-to-Image • Updated Nov 9, 2022 • 2.65k
- tiiuae/falcon-40b**
Text Generation • Updated Sep 29, 2023 • 33.9k • 2.4k
- hakurei/waifu-diffusion**
Text-to-Image • Updated Jul 6, 2023 • 29.9k • 2.36k
- openai/whisper-large-v3**
- stabilityai/stable-video-diffusion-img2vid-xt**

<https://huggingface.co/>

Information Extraction



Examples

- Customer Support Automation:
 - Extracting customer information, issue types, and sentiments from support tickets or emails.
 - Routing support requests to the appropriate department or agent based on extracted information.
 - Generating automated responses or suggestions based on extracted customer intents and issues.
- Contract Analysis:
 - Extracting key entities, such as parties, dates, obligations, and terms, from legal contracts.
 - Identifying relationships between entities, such as rights, responsibilities, and dependencies.
 - Facilitating contract review, compliance checking, and risk assessment processes.
- Financial News Analysis:
 - Extracting company names, financial metrics, and events from financial news articles or reports.
 - Identifying relationships between companies, such as partnerships, mergers, or acquisitions.
 - Monitoring market trends, sentiment analysis, and generating financial insights.
- Healthcare Information Extraction:
 - Extracting patient information, medical conditions, treatments, and medications from clinical notes or medical records.
 - Identifying relationships between symptoms, diagnoses, and treatments for clinical decision support.
 - Facilitating medical research, drug discovery, and patient cohort identification.
- Social Media Monitoring:
 - Extracting mentions of brands, products, or competitors from social media posts or reviews.
 - Identifying sentiment, opinions, and user experiences related to specific entities.
 - Monitoring brand reputation, tracking customer feedback, and identifying influencers or trends.

Examples

- **Résumé and Job Posting Matching:**
 - Extracting skills, qualifications, and experience from job applicants' résumés.
 - Identifying key requirements and qualifications from job postings.
 - Matching job applicants to relevant job openings based on extracted information.
- **E-commerce Product Categorization:**
 - Extracting product attributes, specifications, and categories from product descriptions or catalogs.
 - Identifying relationships between products, such as compatibility, accessories, or variations.
 - Improving product search, recommendation, and inventory management systems.
- **Fraud Detection:**
 - Extracting entities and relationships from transaction records or customer data.
 - Identifying patterns, anomalies, or suspicious activities based on extracted information.
 - Enhancing fraud detection models and risk assessment processes.
- **Research Literature Analysis:**
 - Extracting key entities, such as authors, institutions, and research topics, from scientific publications.
 - Identifying relationships between entities, such as citations, collaborations, or research trends.
 - Facilitating literature reviews, knowledge discovery, and research trend analysis.
- **News Media Monitoring:**
 - Extracting named entities, such as persons, organizations, and locations, from news articles.
 - Identifying events, relationships, and sentiments associated with specific entities.
 - Monitoring media coverage, tracking public opinion, and generating news summaries or alerts.

A large, fluffy white cloud is centered in the upper half of the image. A silver ladder is positioned vertically below the cloud, extending from the bottom edge of the frame towards the base of the cloud. The background is a solid, clear blue sky. The text "Topic Modelling" is written in a bold, black, sans-serif font across the middle of the cloud.

Topic Modelling

Examples

- Content Categorization and Tagging:
 - Automatically categorizing articles, blog posts, or documents into predefined topics or categories.
 - Generating tags or keywords for content based on the identified topics.
 - Improving content organization, searchability, and recommendation systems.
- Customer Feedback Analysis:
 - Identifying common themes or topics in customer reviews, surveys, or feedback.
 - Analyzing sentiment associated with each topic to gauge customer satisfaction.
 - Discovering areas for improvement, product enhancements, or customer service optimization.
- Social Media Monitoring:
 - Detecting trending topics or conversations on social media platforms.
 - Identifying user interests, preferences, and opinions based on the topics they engage with.
 - Tailoring marketing strategies, content creation, and user engagement based on identified topics.
- Research Literature Analysis:
 - Discovering main research themes or areas within a large corpus of scientific publications.
 - Identifying relationships between research topics, authors, or institutions.
 - Facilitating literature reviews, research trend analysis, and knowledge discovery.
- News Media Analysis:
 - Identifying major news topics, events, or stories across multiple news sources.
 - Tracking the evolution of news topics over time and detecting emerging trends.
 - Generating news summaries or alerts based on specific topics of interest.

Examples

- E-commerce Product Analysis:
 - Identifying common themes or features in product reviews or customer feedback.
 - Discovering product categories or segments based on topic similarities.
 - Improving product recommendations, search relevance, and customer segmentation.
- HR and Talent Management:
 - Analyzing employee feedback, surveys, or performance reviews to identify common topics or concerns.
 - Discovering skills, competencies, or qualities mentioned in job descriptions or résumés.
 - Facilitating talent acquisition, employee engagement, and workforce planning strategies.
- Legal Document Analysis:
 - Identifying key topics or themes in legal contracts, agreements, or case documents.
 - Discovering relationships between legal topics, parties, or clauses.
 - Enhancing legal research, contract review, and compliance analysis processes.
- Healthcare and Medical Research:
 - Identifying prevalent topics or themes in patient feedback, medical records, or research papers.
 - Discovering patterns, trends, or associations between medical conditions, treatments, or outcomes.
 - Supporting clinical decision-making, drug discovery, and personalized medicine initiatives.
- Educational Content Analysis:
 - Identifying main topics or concepts covered in educational materials, such as textbooks, lecture notes, or online courses.
 - Discovering relationships between topics, prerequisites, or learning objectives.
 - Facilitating curriculum development, content recommendation, and personalized learning experiences.