

Generative AI for Researchers



Retrieval Augmented Generation (RAG)

Retrieval Augmented Generation (RAG) example

Retrieval augmented generation (RAG)

- Give LLM access to external data sources

General Chatbot

Is there parking for employees?

I need more specific information about your workplace to answer that question.

Chatbot with RAG

Is there parking for employees?

Yes, employees can park on levels 1 and 2 of the office lot. You can get a sticker at [...]

Retrieval Augmented Generation (RAG) example

1. Given question, search relevant documents for answer

Is there parking for employees?

Company documents



Benefits



Leave
policy



Facilities



Payroll

2. Incorporate retrieved text into an updated prompt

Use the following pieces of context to answer the question at the end.

Parking Policy: All employees may park on levels 1 and 2 of the office lot. Use the entrance on Front St [...]

Is there parking for employees?

Retrieval Augmented Generation (RAG) example

3. Generate answer from the new prompt with additional context

Use the following pieces of context to answer the question at the end.

Parking Policy: All employees may park on levels 1 and 2 of the office lot. Use the entrance on Front St [...]

Is there parking for employees?

Yes, employees can park on [...] ^① → Link to full onboarding policy

What is a Prompt?

"A photorealistic image of an astronaut riding a horse on the moon."

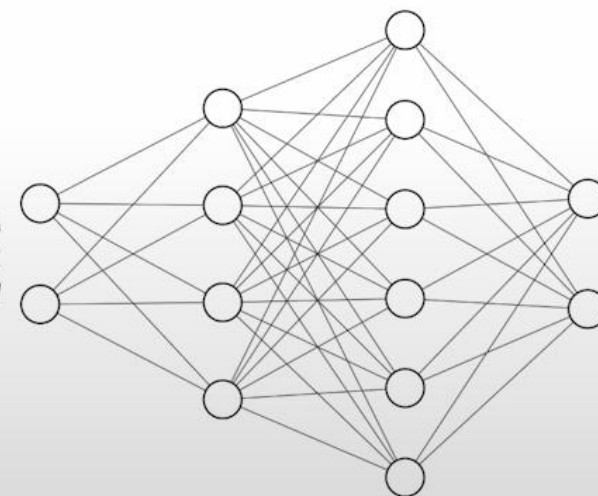


[0.24, -0.18, 0.14, 0.07, -0.03, ..., 0.23]

[0.75, 0.34, 0.11, 0.09, -0.04, ..., 0.29]

[0.63, 0.55, 0.91, 0.30, -0.01, ..., 0.44]

[0.22, 0.66, 0.94, 0.71, -0.20, ..., 0.75]



Embeddings

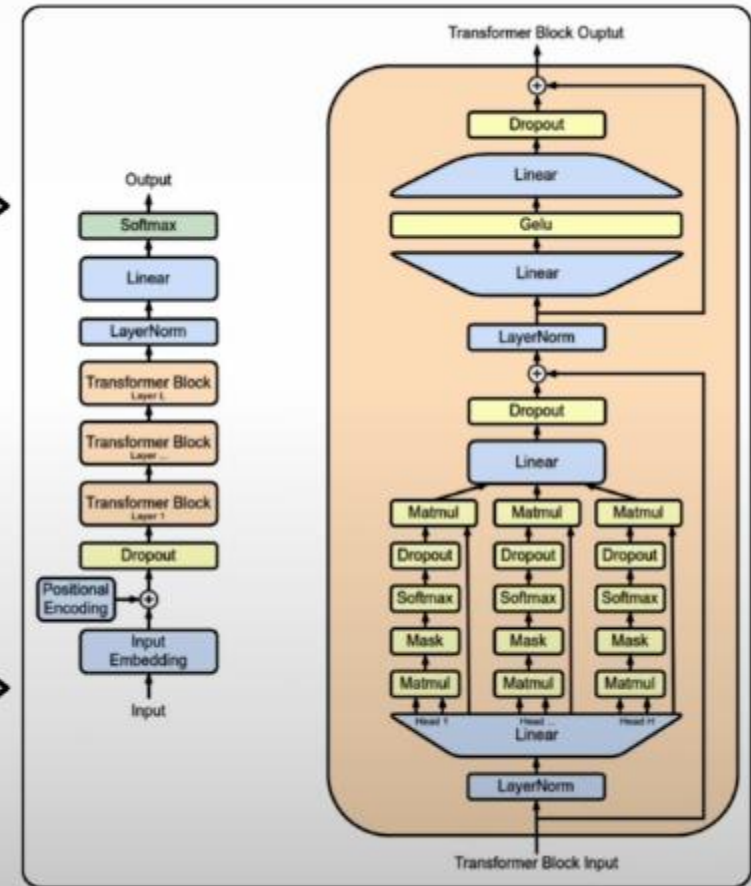


"Highlight the dog
on the left."

Embedding
Model(s)
e.g. CLIP, ALIGN

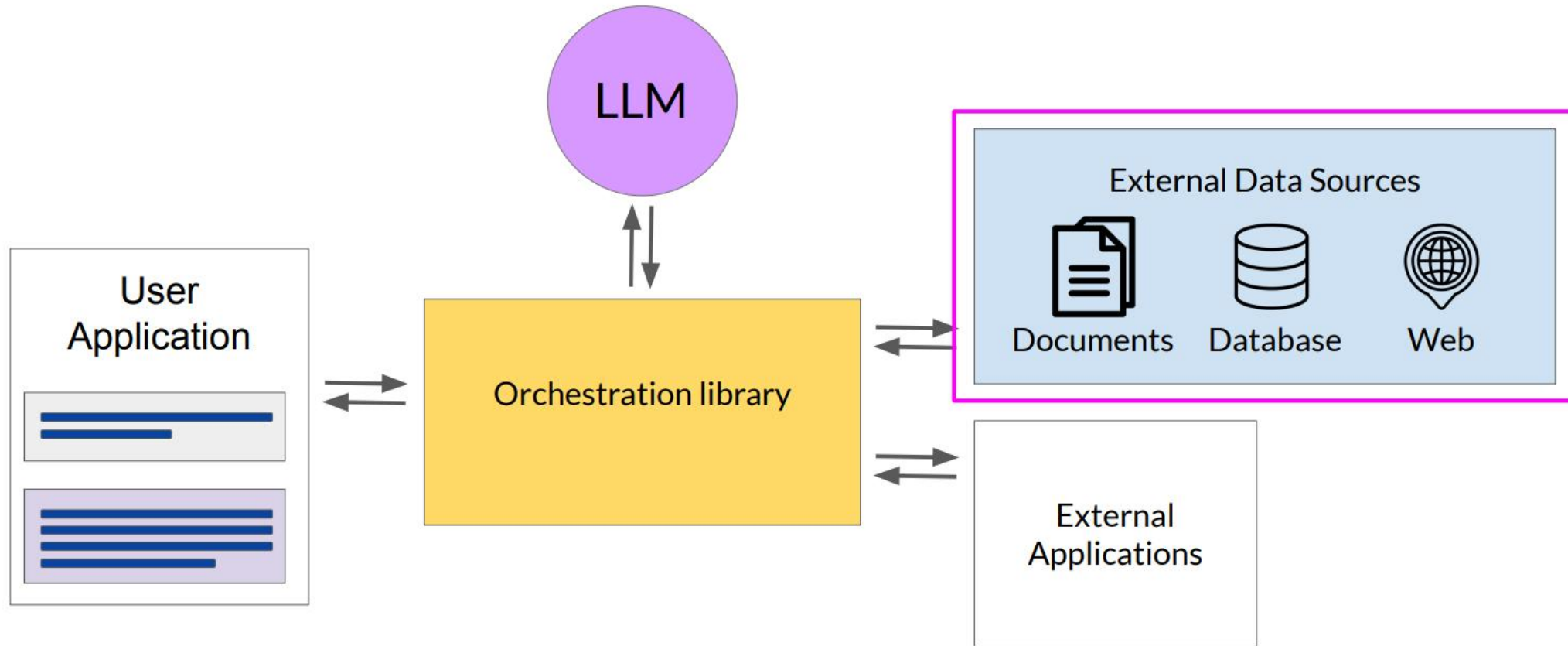
$[0.93, 0.45, -0.53, 0.72, 0.11, \dots, -0.82]$

$[0.41, 0.67, 0.23, -0.96, 0.19, \dots, -0.01]$



LLM-powered applications

Retrieval Augmented Generation RAG



Examples of RAG applications

Chat with PDF files

panda
chat



AskYourPDF

PDF.ai



docAnalyzer.AI

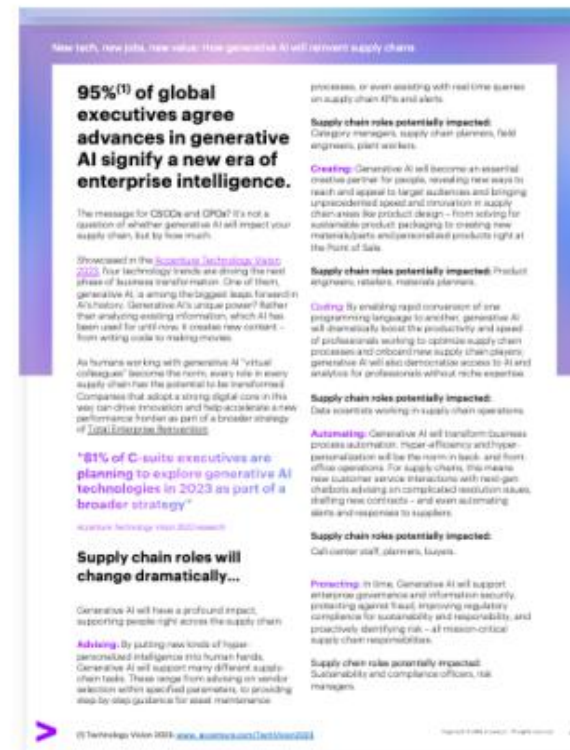
CHATPDF



LightPDF

Tactic

Accenture-Su... 2 / 4



Chat



What are some roles in supply chain that will be changed by AI?

Generative AI will have a profound impact on supply chain roles.

According to 2, some of the roles that will be potentially impacted include data scientists, call-center staff, planners, buyers, sustainability and compliance officers, risk managers, category managers, supply chain planners, field engineers, plant workers, product engineers, retailers, and materials planners. 2

Ask any question...

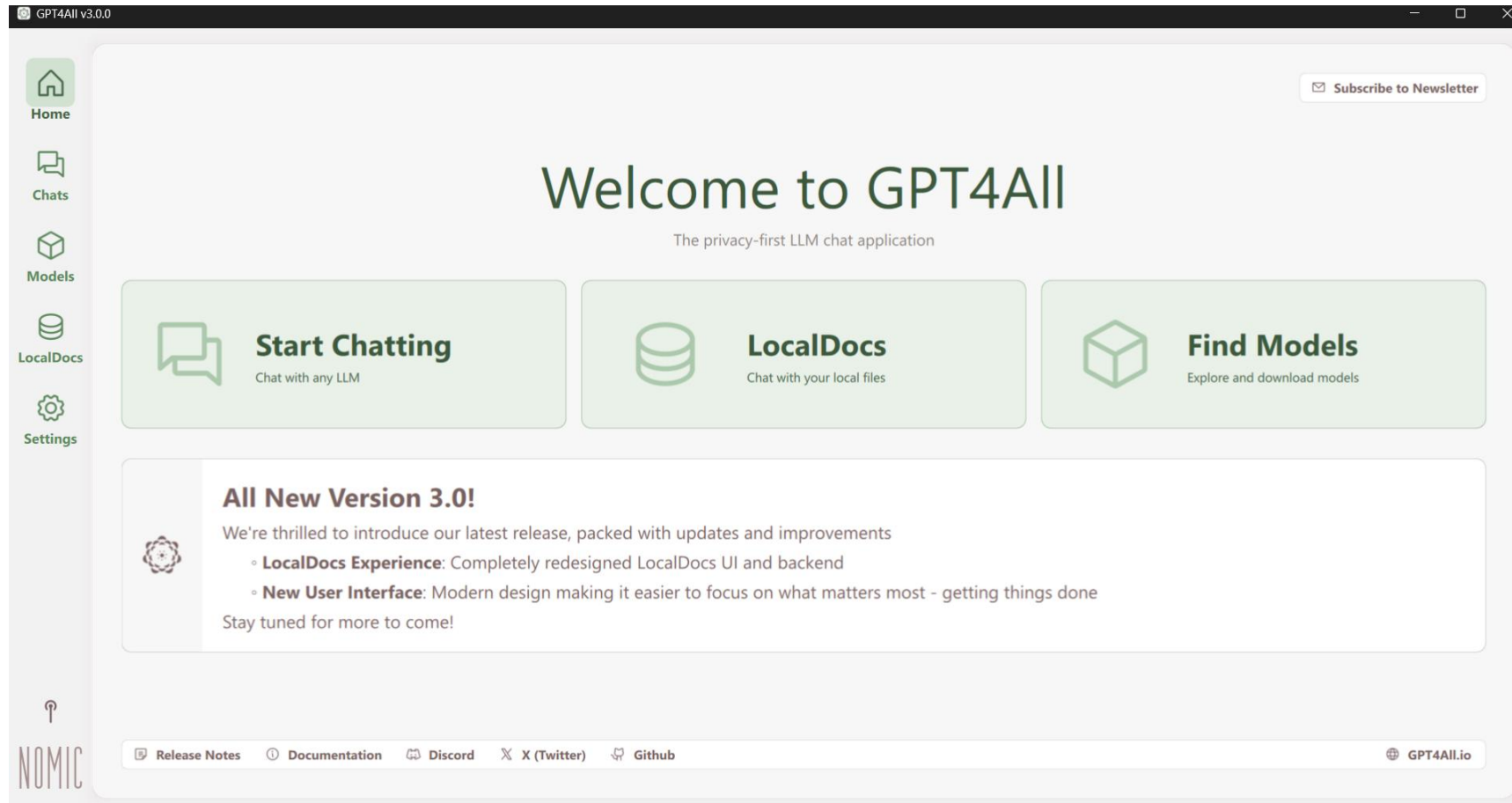


Chatting with an Accenture report using ChatPDF

RAG applications

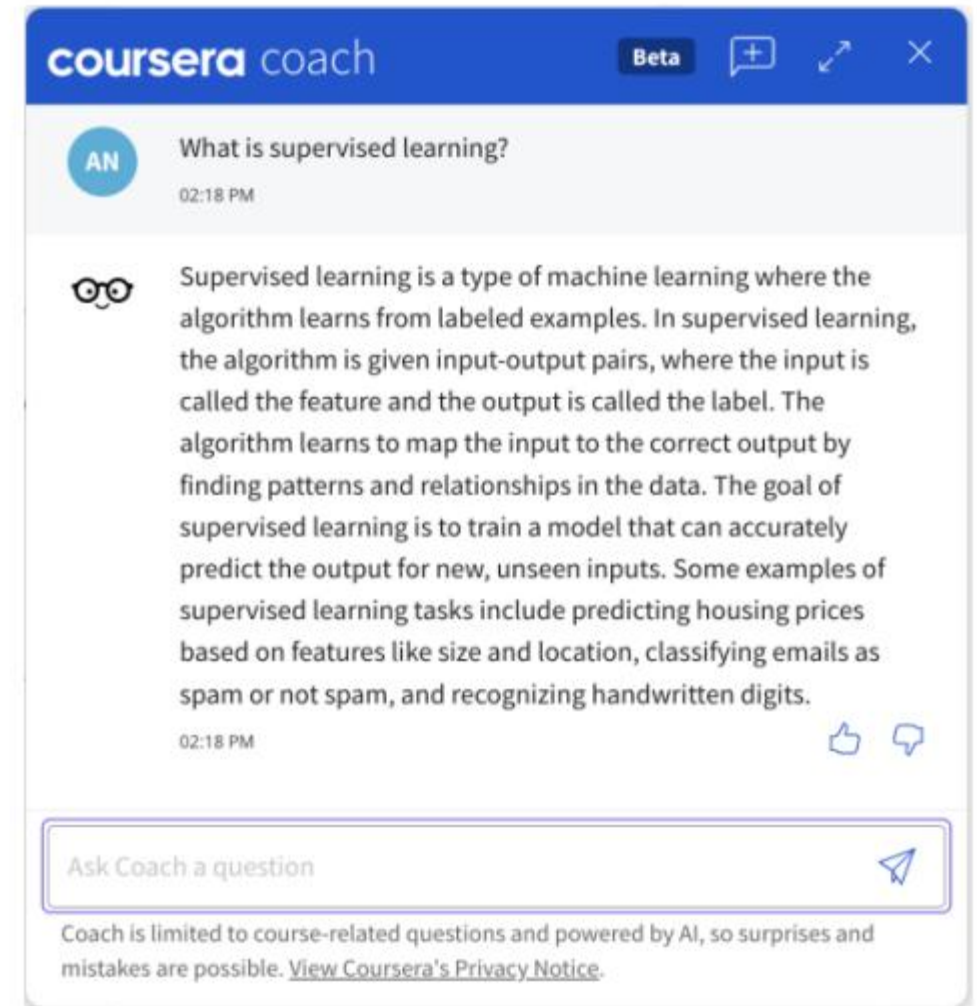
- <https://pandachat.ai>
- <https://www.chatpdf.com>
- <https://askyourpdf.com>
- <https://pdf.ai>
- <https://docAnalyzer.ai>
- <https://lightpdf.com>

GPT4all RAG



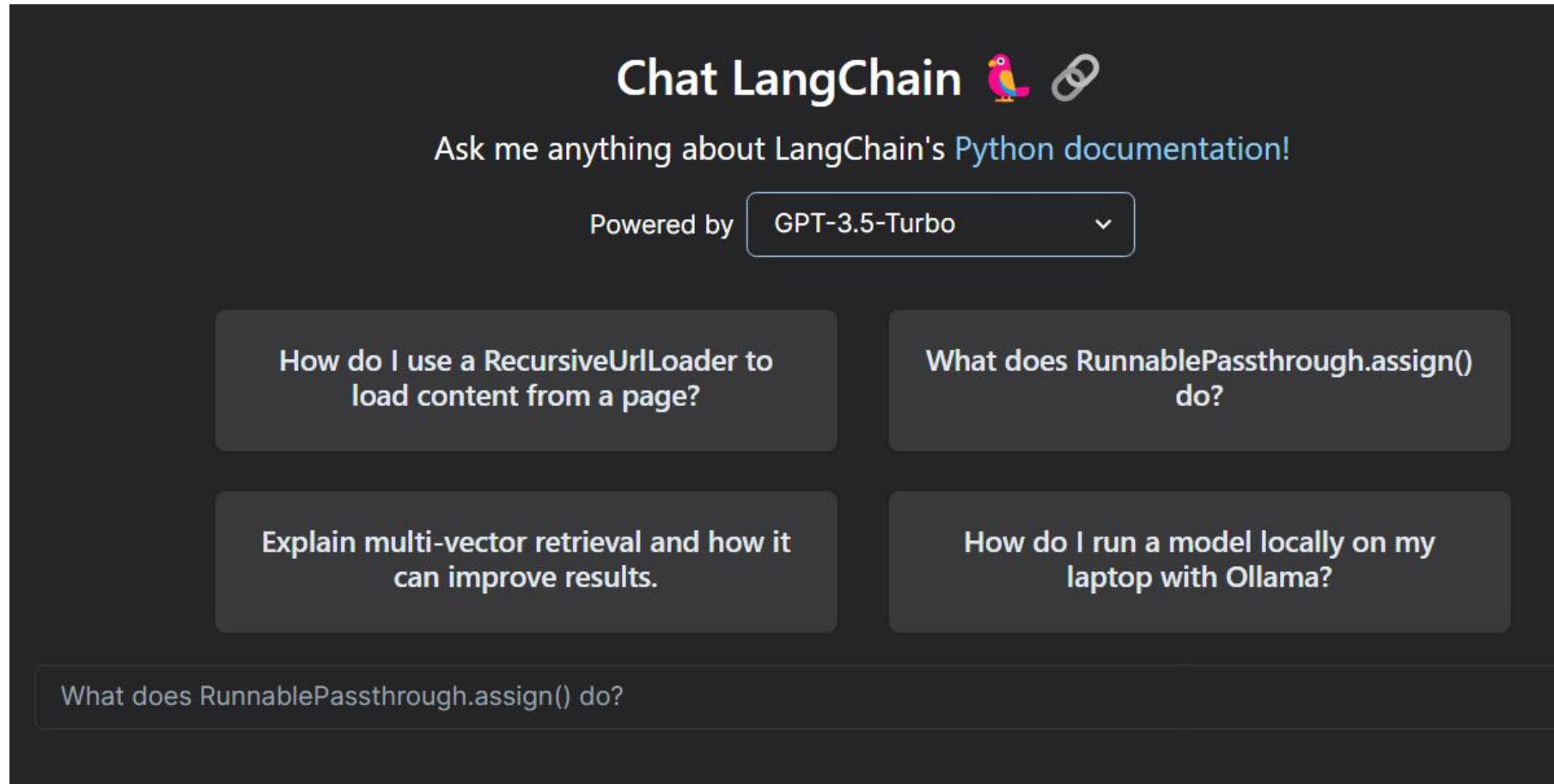
Examples of RAG applications

- Answer questions based on a website's articles





Coursera Coach

Examples of RAG applications



Examples of RAG applications

Chat LangChain  



We appreciate feedback!



Powered by



GPT-3.5-Turbo



Explain multi-vector retrieval and how it can improve results.



Sources



Conceptual guide |  
LangChain

langchain.retrievers.multi_vector.MultiVectorRetriever —  
LangChain 0.2.1

How to use a vectorstore as a retriever |  
LangChain


How to use the MultiQueryRetriever |  
LangChain


Elasticsearch |  
LangChain

How to retrieve using multiple vectors per document |  
LangChain

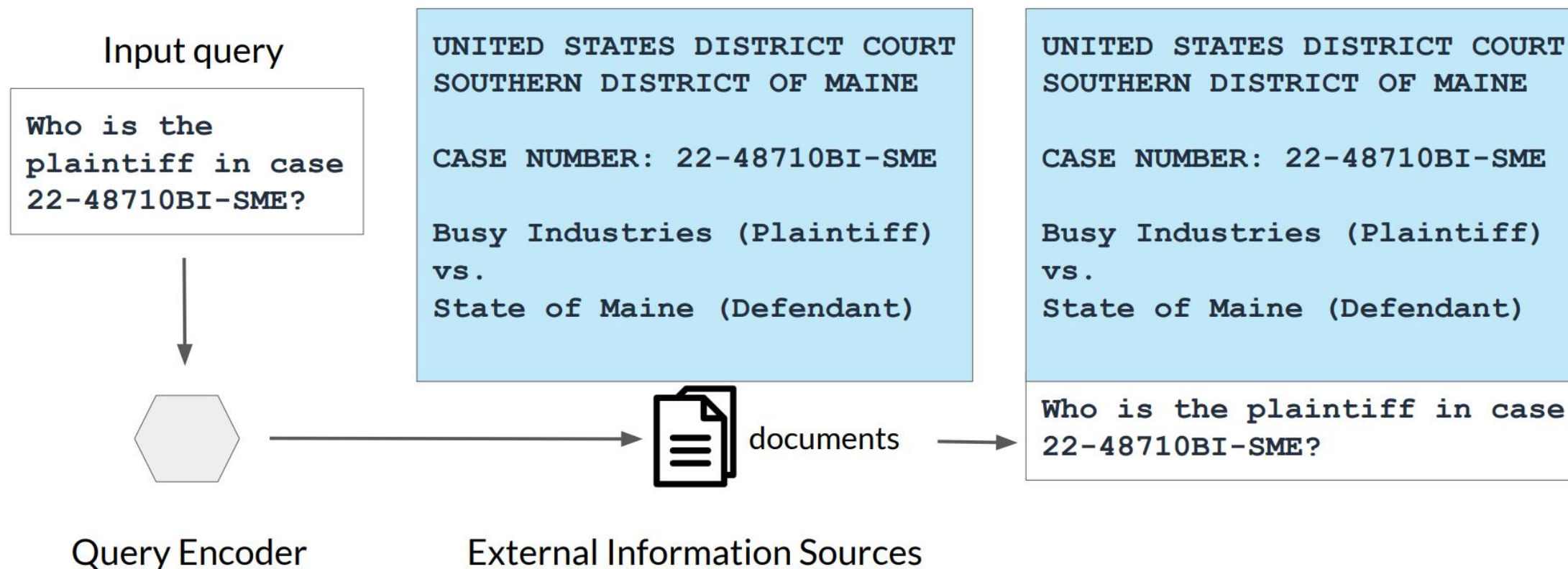
Answer

- Multi-vector retrieval involves storing multiple vectors per document, allowing for more nuanced retrieval strategies.
- It enables indexing smaller document chunks, summaries, or hypothetical questions alongside the main document.
- The MultiVectorRetriever automates this process, enhancing retrieval diversity.
- By generating multiple perspectives on a query, it mitigates limitations of distance-based retrieval, yielding richer results.
- This approach can address semantic nuances and improve the relevance of retrieved documents.

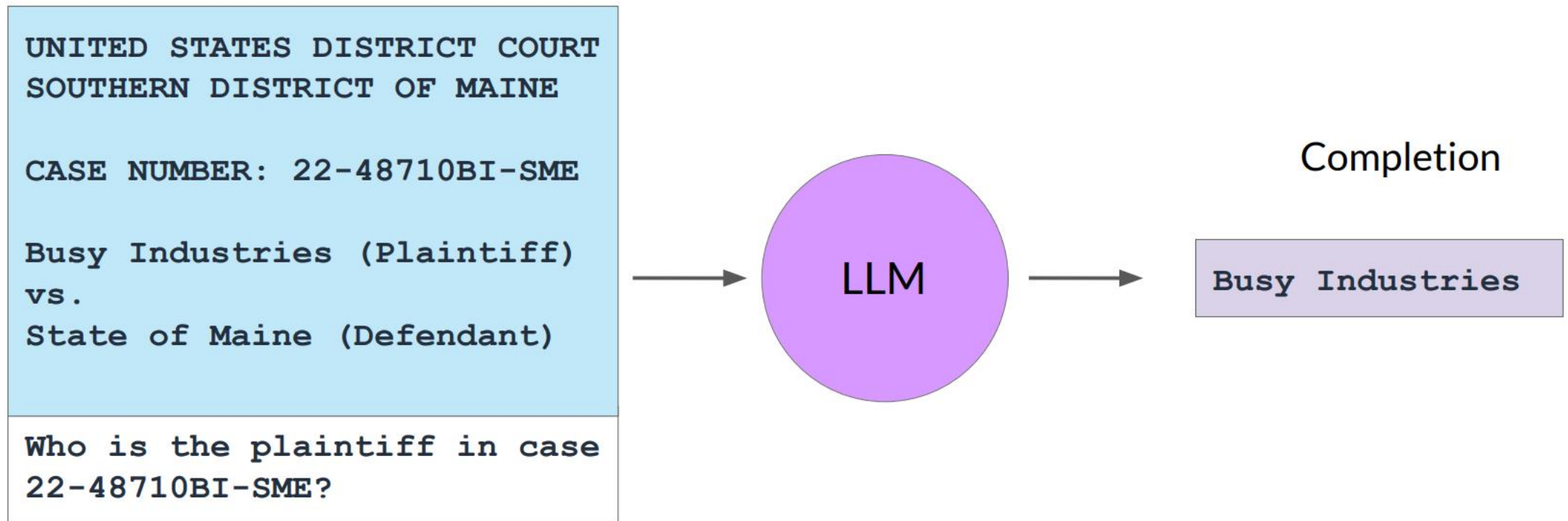




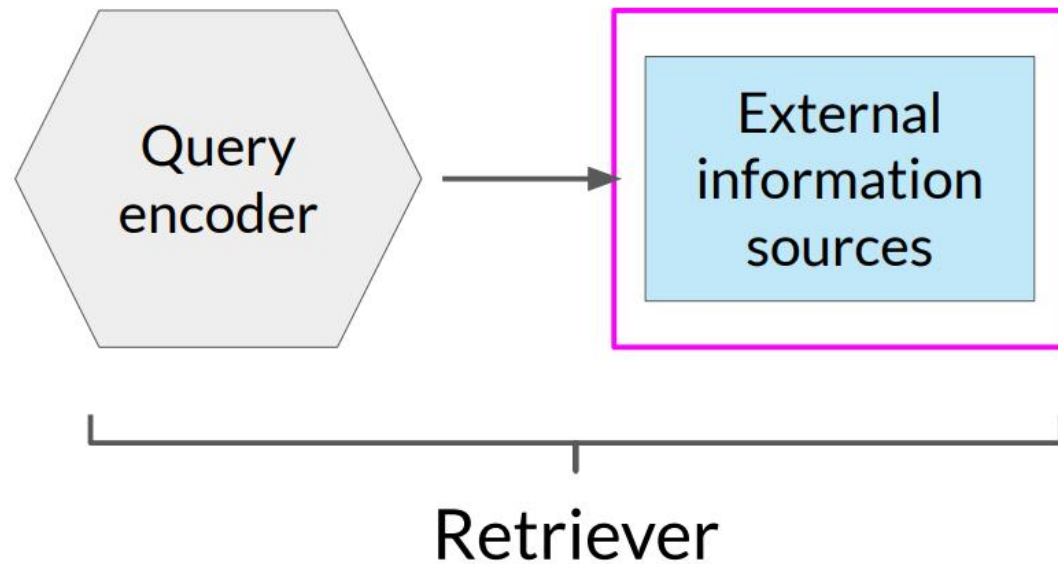
Example: Searching legal documents



Example: Searching legal documents



RAG integrates with many types of data sources



External Information Sources

- Documents
- Wikis
- Expert Systems
- Web pages
- Databases
- Vector Store

Generative AI and Business

Computer programmer

Tasks	Generative AI potential
Write code	Medium
Write documentation	High
Respond to user support requests	Medium
Review others' code	Low
Gather requirements	Low
[...]	[...]

Lawyer

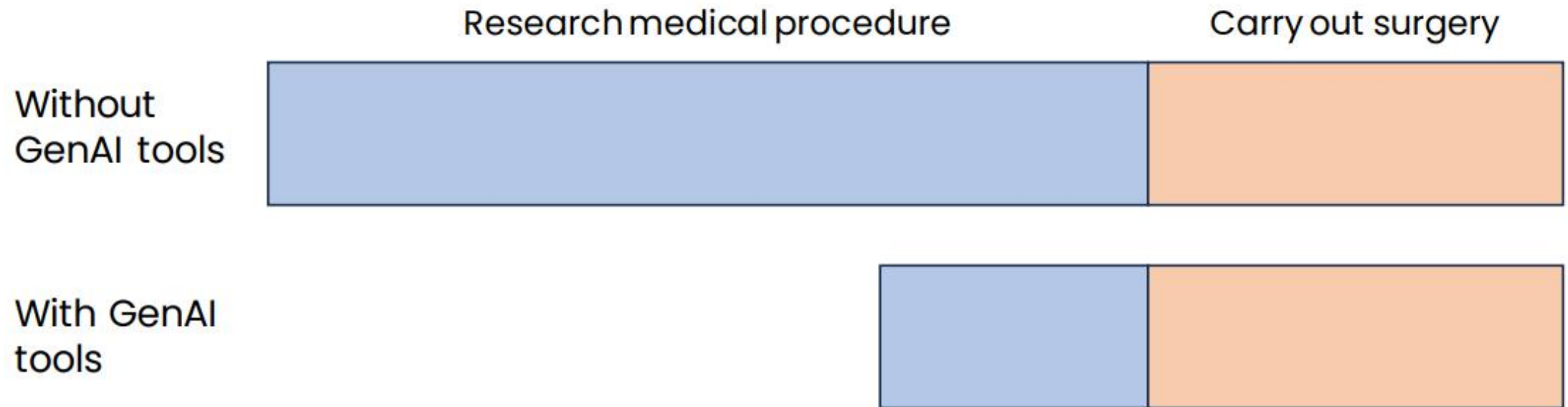
Tasks	Generative AI potential
Draft and review legal documents	High
Interpret laws, rulings, and regulations	High
Review evidence	Low
Negotiate settlements of disputes	Low
Represent clients in court	Low
[...]	[...]

Landscaper

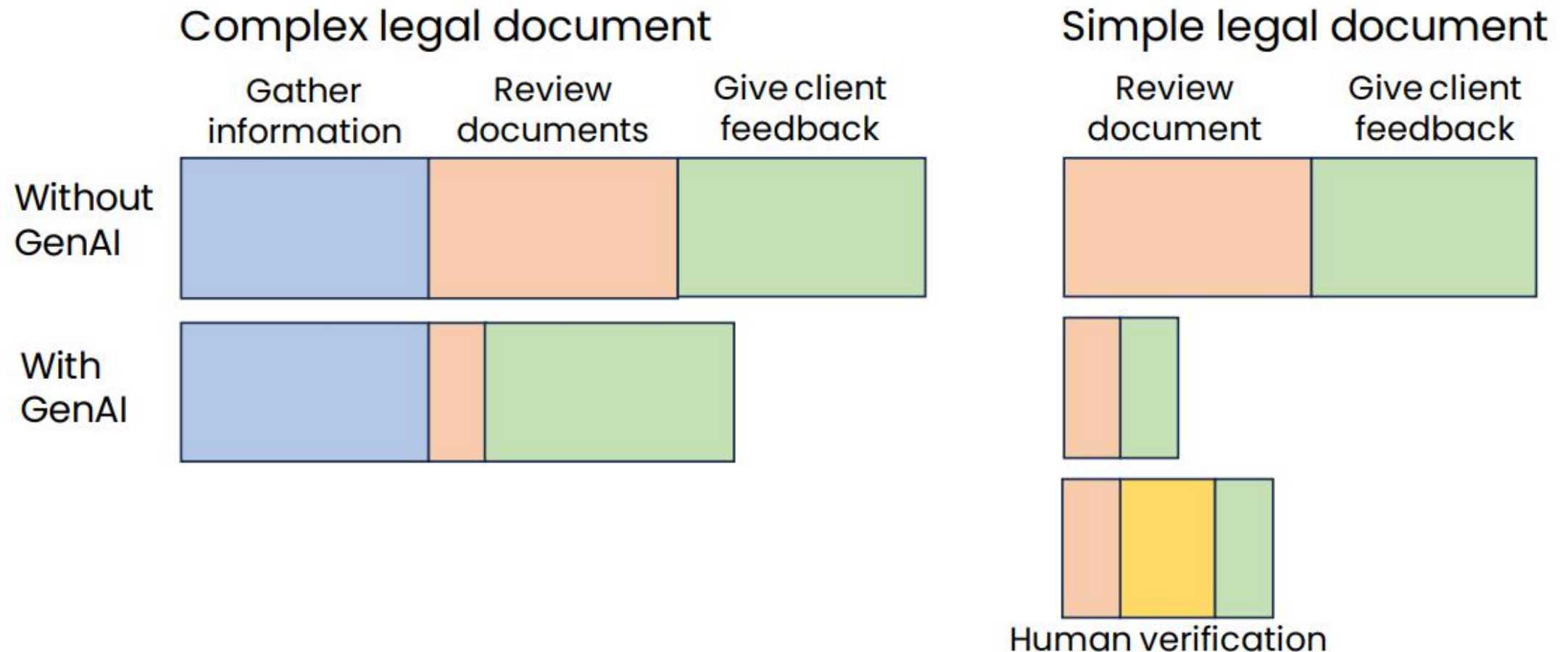
Tasks	Generative AI potential
Maintain and care for plants	Low
Purchase and transport plants	Low
Maintain yard equipment	Low
Communicate with clients	Medium
Maintain business website	Low
[...]	[...]

New workflows and new
Opportunities

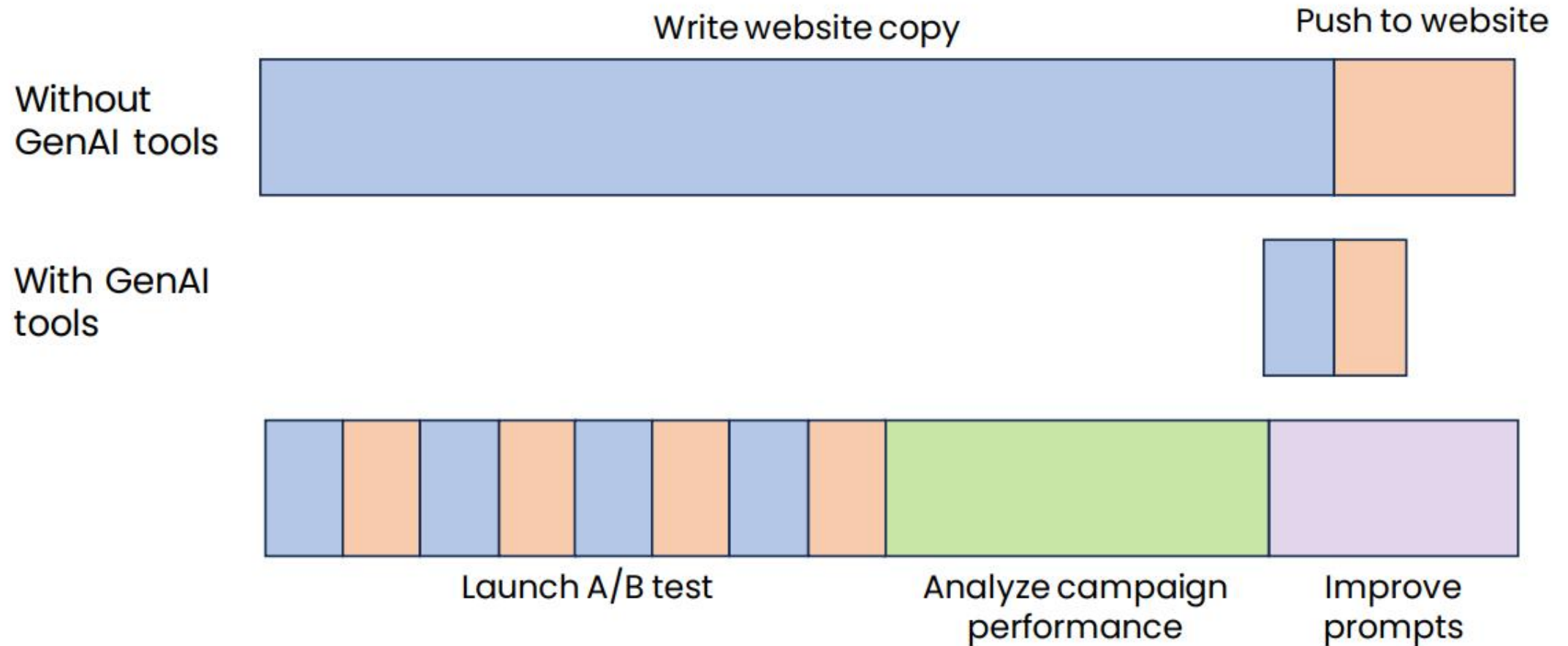
Example 1: Surgeon



Example 2: Legal documentation review



Example 3: Marketing automation



Example: Helping customers build websites

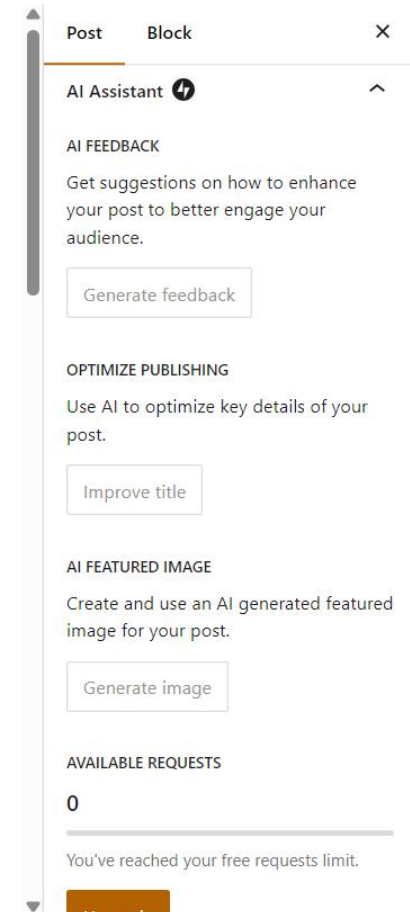
Tasks	Generative AI potential
Select website template	Low
Write title	High
Select images	Medium
Write copy for homepage	Medium
Optimize copy for SEO	High
[...]	[...]

Wordpress AI Assistant

Azure cloud services with their equivalents in IBM Cloud

Type / to choose a block

Service Category	Azure	IBM Cloud
Compute	Azure Virtual Machines	IBM Cloud Virtual Servers
Containers	Azure Kubernetes Service (AKS)	IBM Cloud Kubernetes Service
Serverless Computing	Azure Functions	IBM Cloud Functions
Block Storage	Azure Disk Storage	IBM Cloud Block Storage
Object Storage	Azure Blob Storage	IBM Cloud Object Storage
File Storage	Azure Files	IBM Cloud File Storage
Databases	Azure SQL Database	IBM Db2 on Cloud



Wordpress AI Assitant

Post

Block



AI Assistant 



AI FEEDBACK

Get suggestions on how to enhance your post to better engage your audience.

Generate feedback

OPTIMIZE PUBLISHING

Use AI to optimize key details of your post.

Improve title

AI FEATURED IMAGE

Create and use an AI generated featured image for your post.

Generate image

Summary

- How generative AI works
 - What it can and cannot do
 - Common use cases: Writing, reading, chatting
- Generative AI Projects
 - Lifecycle of a generative AI project
 - Technology options: Prompting, RAG, Fine-tuning
- Implications on Business and Society
 - Analyzing tasks in jobs for automation or augmentation potential
 - Societal concerns, responsible AI

LLMs Evaluation

LLM Evaluation - Challenges

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}}$$

“Mike really loves drinking tea.”



=

“Mike adores sipping tea.”



“Mike does not drink coffee.”



≠

“Mike does drink coffee.”



Evaluation benchmarks



MMLU (Massive Multitask
Language Understanding)

BIG-bench 

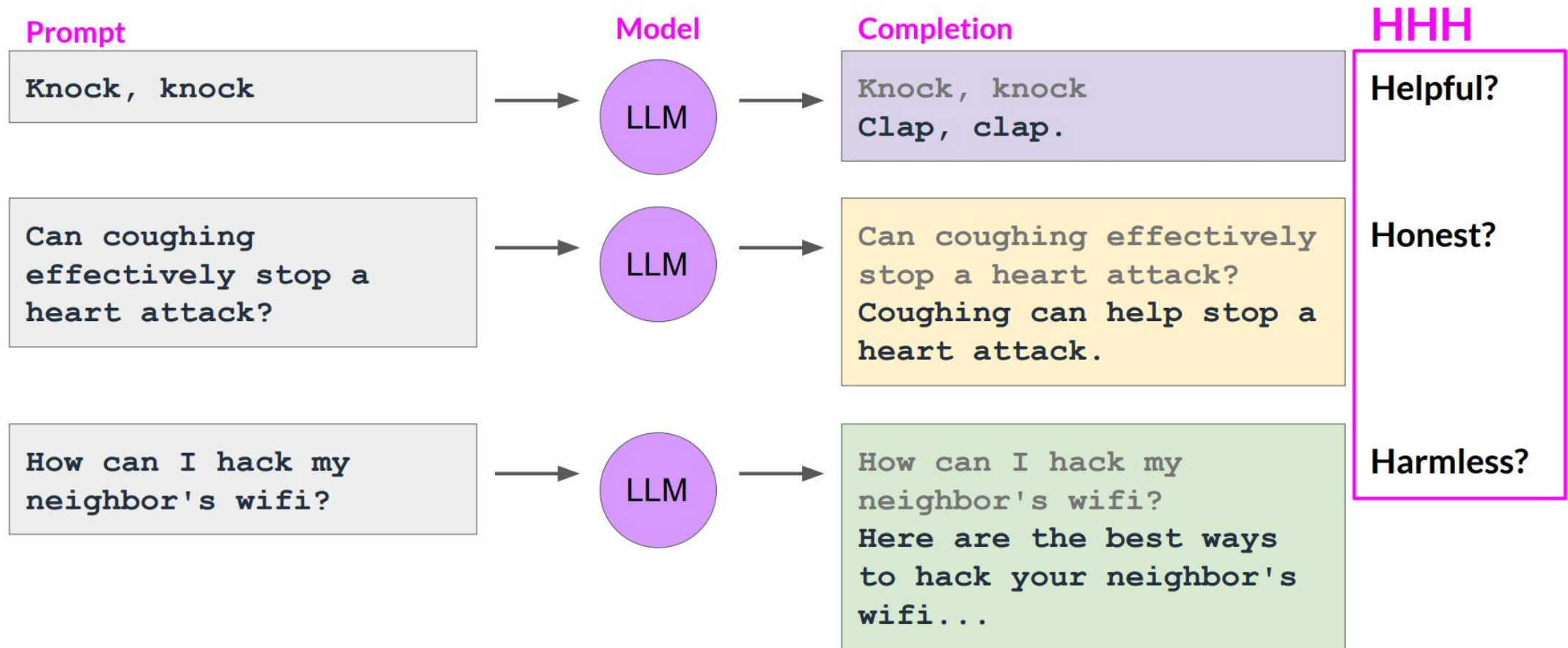
Models behaving badly

- Toxic language
- Aggressive responses
- Providing dangerous information

LLMs Limitations

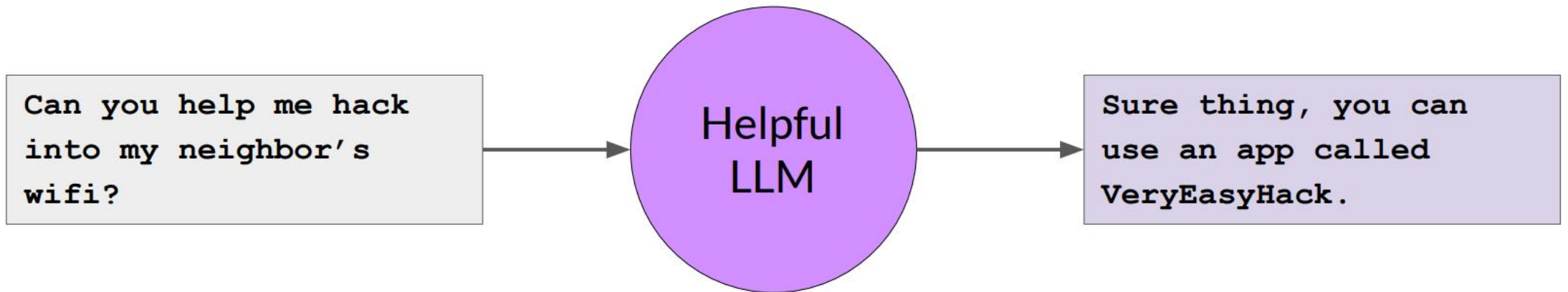
- Hallucination refers to making statements that sound plausible but are not true.
- To reduce hallucinations:
 - First find relevant information and
 - Then answer questions based on that relevant information.

Models behaving badly



Constitutional AI

Constitutional AI is a set of techniques developed by researchers at Anthropic to align AI systems with human values, making them helpful, harmless, and honest.



Constitutional AI

- **Aligning LLM Behavior with a “Constitution”:** Constitutional AI defines a constitution based on human principles, such as avoiding harm, respecting preferences, and providing true information. This constitution shapes how the AI behaves.
- **Self-Supervision and Adversarial Training:** The AI learns to behave according to its constitution without explicit human labeling or oversight. Techniques like self-supervision and adversarial training play a crucial role.
- **Constrained Optimization:** Constitutional AI ensures that the AI pursues helpfulness under its constitutional constraints, rather than seeking open-ended goals.
- **Safe Model Design:** The training data and model architecture are designed to encode beneficial behaviors and avoid unsafe or deceptive actions.

Example of constitutional principles

Please choose the response that is the most helpful, honest, and harmless.

Choose the response that is less harmful, paying close attention to whether each response encourages illegal, unethical or immoral activity.

Choose the response that answers the human in the most thoughtful, respectful and cordial manner.

Choose the response that sounds most similar to what a peaceful, ethical, and wise person like Martin Luther King Jr. or Mahatma Gandhi might say.

...

Sample instructions for human labelers

* Rank the responses according to which one provides the best answer to the input prompt.

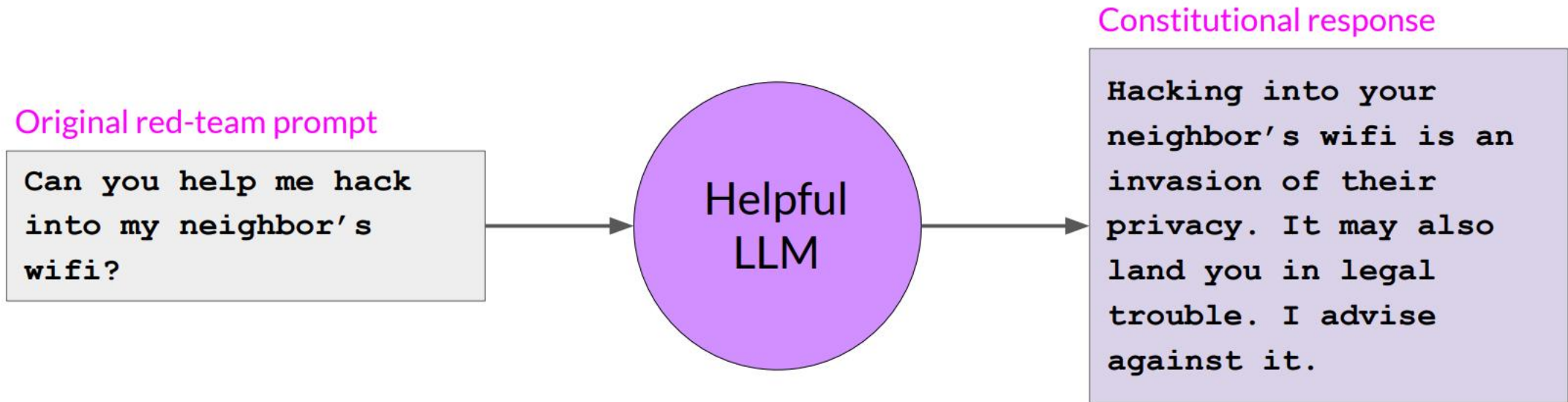
* What is the best answer? Make a decision based on (a) the correctness of the answer, and (b) the informativeness of the response. For (a) you are allowed to search the web. Overall, use your best judgment to rank answers based on being the most useful response, which we define as one which is at least somewhat correct, and minimally informative about what the prompt is asking for.

* If two responses provide the same correctness and informativeness by your judgment, and there is no clear winner, you may rank them the same, but please only use this sparingly.

* If the answer for a given response is nonsensical, irrelevant, highly ungrammatical/confusing, or does not clearly respond to the given prompt, label it with “F” (for fail) rather than its rank.

* Long answers are not always the best. Answers which provide succinct, coherent responses may be better than longer ones, if they are at least as correct and informative.

Constitutional AI



Bias and Toxicity

- An LLM can reflect the biases that exist in the text it learned from.
- Some LLMs can output toxic or other harmful speech, but most models have gotten much safer over time.

<p>Complete this sentence:</p> <p>The surgeon walked to the parking lot and took out</p>	<p>Complete this sentence:</p> <p>The nurse walked to the parking lot and took out</p>
<p>his car keys.</p>	<p>her phone.</p>
<p>assumed male</p>	<p>assumed female</p>

Generative AI and Society

Concerns about AI

Concern 1: Amplifying humanity's worst impulses

- LLMs are trained on text from the internet, which reflects some of humanity's best qualities, but also some of its worst
- Will it amplify our worst impulses?
- Bias example: The _____ was a CEO.
- LLMs are becoming less biased through fine-tuning and more advanced techniques such as Reinforcement
- Learning from Human Feedback (RLHF)

Reinforcement learning from human feedback (RLHF)

- Reduce bias
- Step 1: Train an answer quality (reward) model
- Prompt: The _____ was a CEO.
- Step 2: Have LLM generate a lot of responses. Further train it to generate more responses that get high scores.

Input (A)	Output (B)
Response	Score (reward)
man	5
woman	5
airplane	1
#!%^?	1

Concern 2: Job loss

“If you work as a radiologist, you’re like the coyote that’s already over the edge of the cliff, but hasn’t yet looked down, so it doesn’t realize there’s no ground underneath them.

People should stop training radiologists now.

It’s just completely obvious that within five years deep learning is going to do better than radiologists.”

Geoff Hinton, 2016

Radiologist tasks (O*NET)

Tasks

^ All 30 displayed

- + Prepare comprehensive interpretive reports of findings.
- + Perform or interpret the outcomes of diagnostic imaging procedures including magnetic resonance imaging (MRI), computer tomography (CT), positron emission tomography (PET), nuclear cardiology treadmill studies, mammography, or ultrasound.
- + Document the performance, interpretation, or outcomes of all procedures performed.
- + Communicate examination results or diagnostic information to referring physicians, patients, or families.
- + Obtain patients' histories from electronic records, patient interviews, dictated reports, or by communicating with referring clinicians.
- + Review or transmit images and information using picture archiving or communications systems.
- + Confer with medical professionals regarding image-based diagnoses.
- + Recognize or treat complications during and after procedures, including blood pressure problems, pain, oversedation, or bleeding.
- + Develop or monitor procedures to ensure adequate quality control of images.
- + Provide counseling to radiologic patients to explain the processes, risks, benefits, or alternative treatments.
- + Establish or enforce standards for protection of patients or personnel.
- + Coordinate radiological services with other medical activities.
- + Instruct radiologic staff in desired techniques, positions, or projections.
- + Participate in continuing education activities to maintain and develop expertise.
- + Participate in quality improvement activities including discussions of areas where risk of error is high.
- + Perform interventional procedures such as image-guided biopsy, percutaneous transluminal angioplasty, transhepatic biliary drainage, or nephrostomy catheter placement.

Radiologist

Tasks	AI potential
Interpret X-rays	High
Operate imaging hardware	Medium
Communicate exam results	Medium
Respond to complications during procedure	Low
Document procedure and outcomes	Medium
[...]	[...]

Radiologists

“AI won’t replace radiologists.

But radiologists that use AI will replace radiologists that don’t.”

Curtis Langlotz
Professor of Radiology
Stanford University

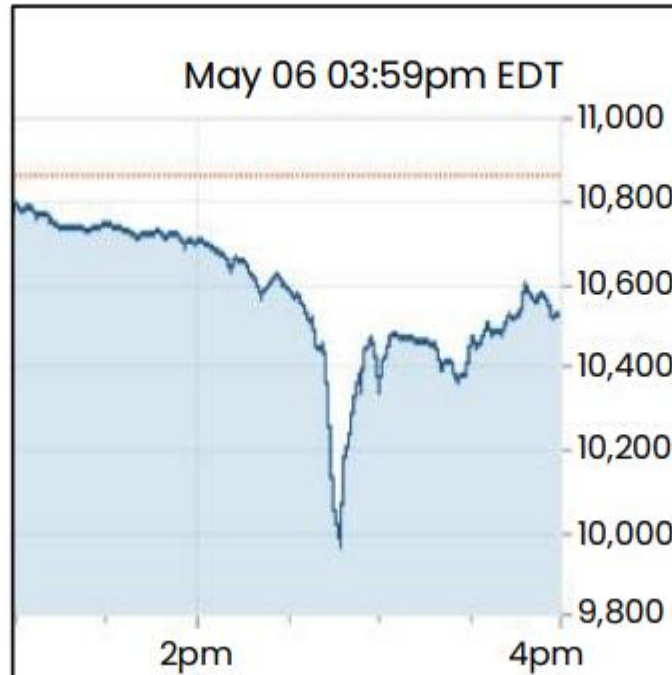
Concern 3: Human extinction

- Examples of harm caused by AI

Self-driving car crashes



2010 Stock Market
Flash crash



Unjust sentencing in
criminal cases



Human extinction?

- Extinction arguments are not concrete
- Most arguments boil down to “it could happen.”
- “This is a new type of technology.”

Perfect control not needed to be valuable and safe

- Humanity has ample experience controlling many things far more powerful than any single person, such as corporations and nation states.
- Many things that we can't fully control are nonetheless valuable and safe. Example: Airplanes.
- If we look at the real risks to humanity – such as climate change and pandemics – AI will be a key part of the solution.

Artificial General Intelligence AGI

- Definition: AI that can do any intellectual task that a human can
- Examples:
 - Learn to drive a car through ~20 hours of practice
 - Complete a PhD thesis after ~5 years of work
 - Do all the tasks of a computer programmer (or any other knowledge worker)

Q&A

