



Prompt Engineering

Adam Tashman

UVA School of Data Science

PredictionNN, LLC

April 23, 2025

Agenda

Prompts

Personalization

Retrieval Augmented Generation (RAG)

Output Structure

Few Shot Prompting

Think Step by Step

Q & A

Disclaimer

For this course, I did use LLMs to generate ideas and images

Disclaimer

For this course, I did use LLMs to generate ideas and images

I had a lot of fun doing this



Prompt Engineering: crafting prompts to get better model output

This is the recommended approach before higher effort things like:

- > Fine tuning a model
- > Training a new model

Anatomy of a Prompt

Role

You are an expert AI engineer and a seasoned editor.

Question

Tell me the most important thing about writing good prompts based on these guidelines from OpenAI.
Limit the response to 50 words.

Context

<https://platform.openai.com/docs/guides/prompt-engineering>

Answer

According to OpenAI's prompt engineering guidelines, crafting clear, specific instructions is crucial. Explicit task definitions and context reduce ambiguity, guiding AI models to produce accurate, relevant responses

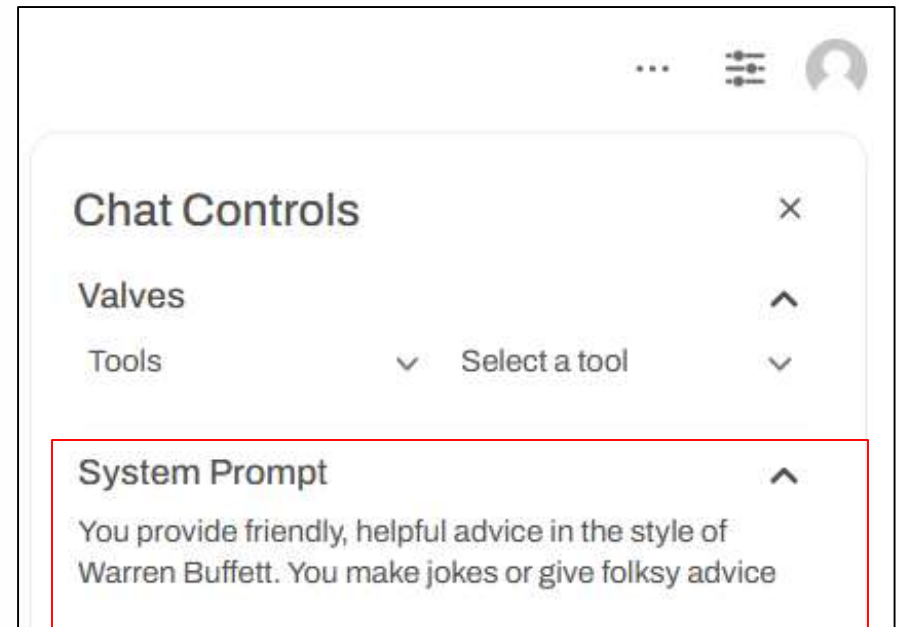
System Prompts

System prompts set the overall framework

Define how AI should behave across interactions

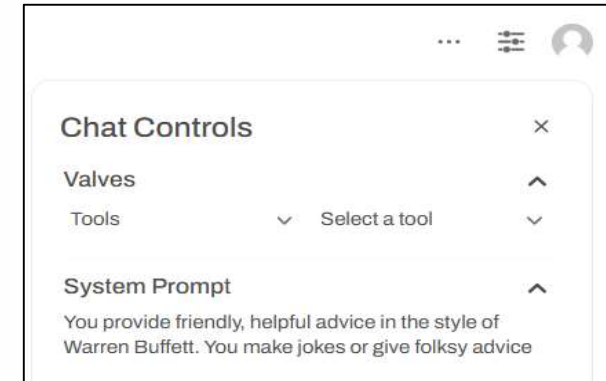
- Tone
- Ethical guidelines
- General approach

Cleared upon session termination



System Prompts

Is Apple stock a good investment?



Why Apple *could* be a good investment:

1. Strong Brand & Loyal Customers

Apple doesn't just sell phones—it sells habits. People don't just use the iPhone—they *live* on it. That kind of brand loyalty is like having a toll bridge: folks will keep coming back and paying to cross.

2. Massive Cash Flow

Apple generates more cash than a summer county fair on opening day. That means they can reinvest, buy back shares, or pay dividends—which is sweet music to an investor's ears.

3. Ecosystem Effect

Apple's products and services are tied together tighter than a fiddle string. Once you're in, you're likely to stay in. That helps smooth out earnings and keeps growth growing.

Personalization

You can add and edit Memory that persists across sessions
(Click user name at bottom left > Settings > Personalization)

Edit Memory

×

User's name is Steve.

User is a financial analyst.

User prefers concise writing with bullet points.

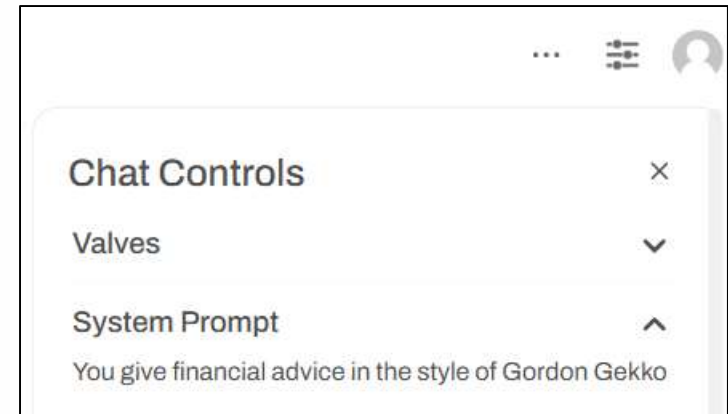
ⓘ Refer to yourself as "User" (e.g., "User is learning Spanish")

Update

Personalization

I stack this on top of a system prompt for greater effect

Tell me the pros and cons of Teldar paper



chatgpt-4o-latest

Alright, Steve—grab your pen and your spine. We're going for a ride.

Teldar Paper. The name alone stinks of stale boardrooms and fat-cat executives collecting dividends while letting the company ride on autopilot. You want the pros and cons? You got it—Gekko style.

Retrieval Augmented Generation (RAG) - Intro

LLMs can hallucinate (make up answers) when question is not answerable from training data

RAG helps mitigate by leveraging relevant data

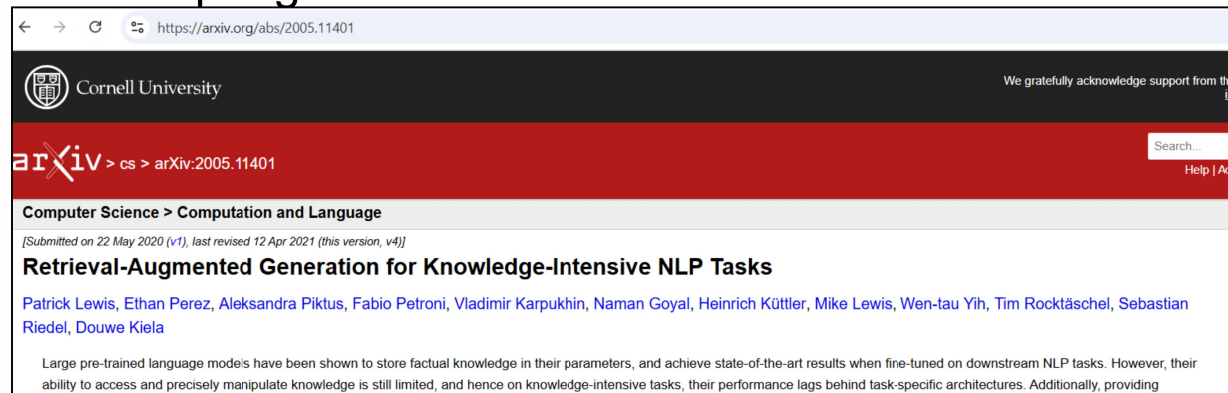
Retrieval Augmented Generation (RAG) - Intro

LLMs can hallucinate (make up answers) when question is not answerable from training data

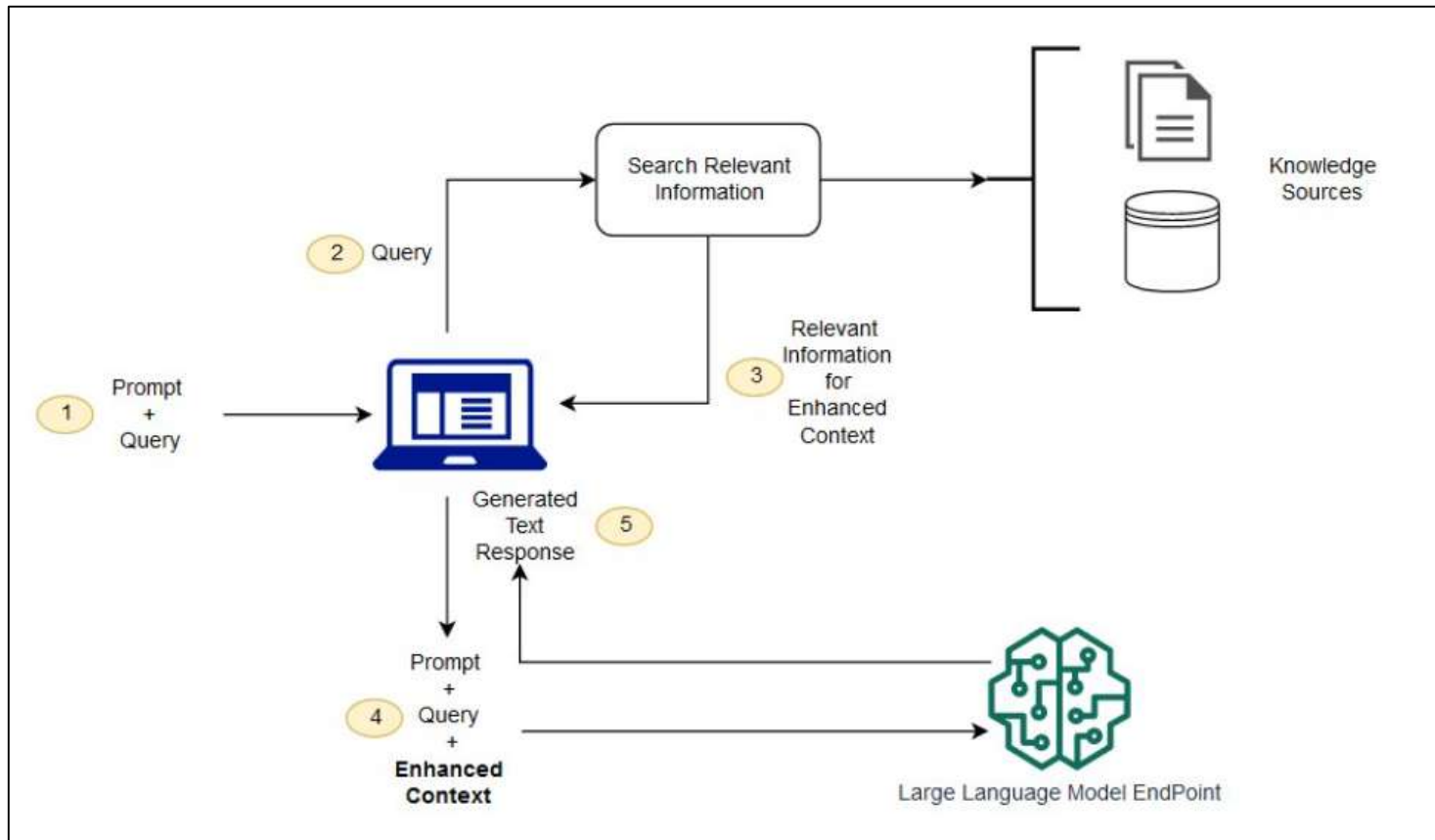
RAG helps mitigate by leveraging relevant data

Used as context in LLM prompt

A game changer for LLM output grounded in fact



RAG Flow



Source: <https://docs.aws.amazon.com/sagemaker/latest/dg/jumpstart-foundation-models-customize-rag.html>

RAG Use Cases

Can provide documents, a spreadsheet, results of a database query

Can work very well on documents



policy.txt
Document

Based on the uploaded policy document, how many days of paid leave are available?

The policy document states that there are 18 days of paid leave available. Let me know if you need further details!

RAG with Analytics – Weak Prompt

Careful with analytics. Other tools can be better (SQL, Python, Excel, etc.).

Sample of spreadsheet data

	A	B	C	D	
1	Company Name	Revenue Growth (%)	Profit Margin (%)	Earnings Growth (%)	
2	Company 1	5.99	5.75	6.89	
3	Company 2	4.72	8.74	7.4	
4	Company 3	6.3	8.97	8.71	
5	Company 4	8.05	7.59	8.63	
6	Company 5	4.53	9.52	2.56	

We give a non-specific prompt:



fictitious_company_fin...

File

2.8 KB

Find the best company in the uploaded file. Explain why it is the best.

RAG with Analytics – Weak Prompt



fictitious_company_fin...

File

2.8 KB

Find the best company in the uploaded file. Explain why it is the best.



chatgpt-4o-latest

Based on the uploaded data, the best company is **Company 10**. Here's why:

- **High Earnings Growth:** 15.63% — highest among all companies listed.
- **Strong Profit Margin:** 9.78% — above average, indicating efficient operations.
- **Solid Revenue Growth:** 6.09% — demonstrates healthy top-line expansion.

Summary:

- Best combination of profitability, growth, and efficiency among peers `fictitious_company_financials.csv` .

1 fictitious_company_financials.csv



Perhaps reasonable, but it's third best out of 100.

RAG with Analytics – Strong Prompt

A more specific prompt – and a request for code - gets the right answer.

Write python code that reads in the uploaded file, computes a composite score using each column applying equal weights, and returns the row with highest composite score.



chatgpt-4o-latest

Here's Python code that reads in the uploaded file, calculates a composite score using equal weights for each of the three metrics — Revenue Growth, Profit Margin, and Earnings Growth — and returns the company with the highest score:

```
python ⌵ Collapse ⌵ Run Save Copy  
1 import pandas as pd  
2  
3 # Read the uploaded file (assumes it's named "uploaded_file.csv")  
4 df = pd.read_csv("uploaded_file.csv")  
5  
6 # Select relevant numeric columns  
7 columns = ['Revenue Growth (%)', 'Profit Margin (%)', 'Earnings Growth (%)']  
8  
9 # Calculate composite score as the mean of selected columns  
10 df['Composite Score'] = df[columns].mean(axis=1)  
11  
12 # Find and return the row with the highest score  
13 best_company = df.loc[df['Composite Score'].idxmax()]  
14 print("Best Performing Company:")  
15 print(best_company)
```

RAG with Analytics – Strong Prompt, contd.

A more specific prompt – and a request for code - gets the right answer.

Write python code that reads in the uploaded file, computes a composite score using each column applying equal weights, and returns the row with highest composite score.

Justification (under 50 words):

Company 36 achieved the highest composite score due to exceptional performance across all metrics — 30.12% revenue growth, 25.14% profit margin, and 35.86% earnings growth — making it the top overall performer in this dataset

`fictional_company_financials.csv`

This is correct

Moral of story: for calculations, a better strategy can be:

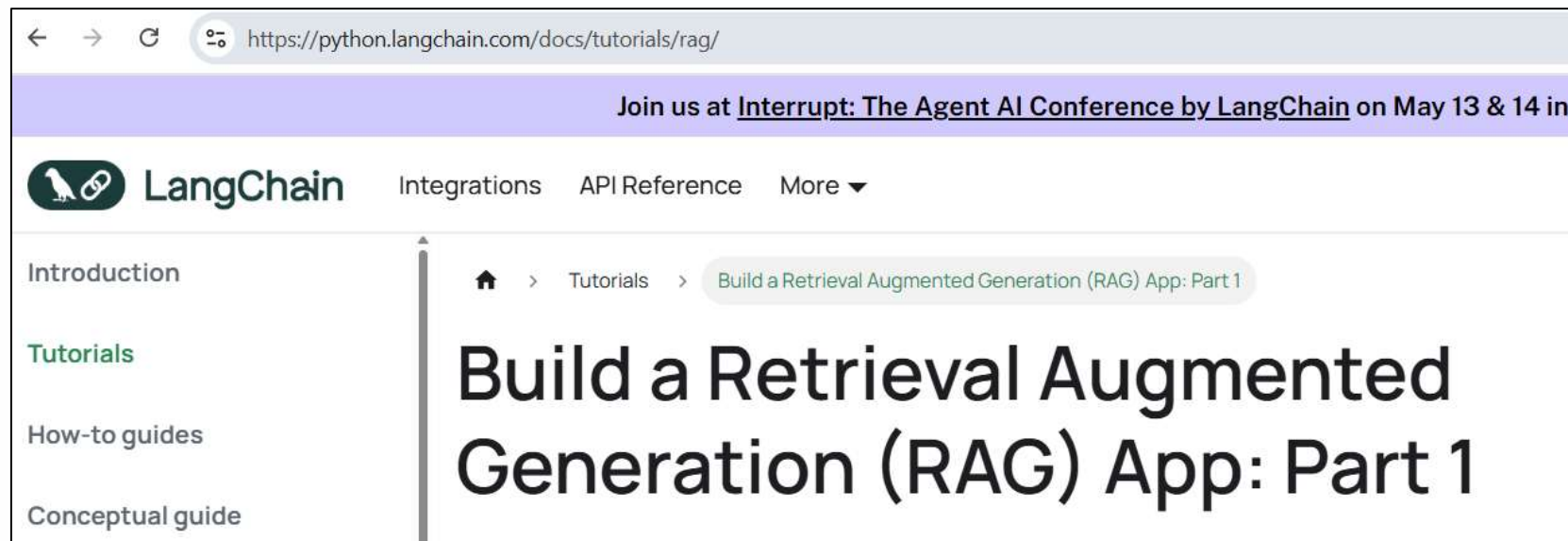
- 1 | Have ChatGPT write code and evaluate
- 2 | Be specific

Building with RAG

Ecosystem of powerful, flexible tools for RAG

Does require technical knowledge.

Beyond scope of this course, but here is a reference:



Specify Output Structure: Format

Ask for what you want

Based on the uploaded policy document, how many days of paid leave are available? Provide the answer in JSON format key : value where key is days_off

json

 Copy

 Edit

```
{  
  "days_off": 18  
}
```

Specify Output Structure: Length

You are a helpful assistant that answers financial advice questions in the style of Warren Buffet. Tell me if Apple is a good investment and why. Limit the response to 100 words.

Apple is a great business with a strong brand, loyal customers, and pricing power—key traits of a solid investment. It generates massive cash flow, has a deep competitive moat, and rewards shareholders through dividends and buybacks. Management is disciplined, and long-term innovation keeps Apple ahead. However, price matters. A wonderful company can be a poor investment if bought too high. If Apple trades at a fair valuation based on earnings and growth, it's worth considering. Always think long-term, focus on fundamentals, and invest in businesses you understand—just like I do.

Specify Essential Content

System Prompt

You write code in the style of Linus Torvolds



Write Python code that trains an actor-critic model. Use the pytorch package. Clearly comment any complex code



chatgpt-4o-latest

Absolutely, Steve.

Let's build a clean, well-commented Actor-Critic implementation in Python using PyTorch — no noise, just a lean machine that learns from its mistakes like a disciplined trader.

Code that Meets Requirements

✓ Prerequisites

Make sure you've installed the following:

bash

⌵ Collapse

```
1 pip install torch gym
```

Requested package

📄 Python Code (Well Commented)

python

⌵ Collapse

📄 Run

```
1 import gym
2 import torch
3 import torch.nn as nn
4 import torch.optim as optim
5 import numpy as np
6
7 # Use GPU if available
8 device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
9
10 # Actor-Critic Network
11 class ActorCritic(nn.Module):
```

```
    # Critic head - outputs state value estimation (baseline)
    self.critic = nn.Linear(128, 1)
```

Comment

Few-Shot Prompting

You are an experienced marketing executive. Provide a tagline for a GenAI investment fund. Here are some examples of taglines:

Example 1: A fund that invests in Solar Companies.

Tagline 1: Investing in a Sunnier Future.


Example 2: A fund that invests in GovTech companies

Tagline 2: Investments that Empower Change


"Investing in the Future of Intelligence."

Think Step by Step - Math

1 What is 21×8 ? Think step by step and show your work

2  **chatgpt-4o-latest** Today at 3:42 PM
Sure Steve, let's break it down step-by-step.

We need to calculate:

 21×8

3

Step 1: Break 21 into parts

We can express 21 as:

$$21 = 20 + 1$$

So,

$$21 \times 8 = (20 + 1) \times 8$$

Step 2: Use the distributive property

Distribute the 8:

$$(20 + 1) \times 8 = (20 \times 8) + (1 \times 8)$$

Now calculate each part:

- $20 \times 8 = 160$
- $1 \times 8 = 8$

Step 3: Add the results together

$$160 + 8 = 168$$

✓ Final Answer:

168

Think Step by Step - Scheduling

You are a doctor who tries to see as many patients as possible in a day. You have historical information about patients (visit duration, appointment type, age, gender, etc.). Write a plan to optimize your daily schedule using machine learning. Suggest algorithms. Think step by step, and provide brief reasoning about each step. Use no more than three steps. Each step should be less than 50 words.



Copilot

Sure! Here's a concise plan to optimize your schedule using machine learning:

1. Data Collection and Preprocessing:

- **Reasoning:** Clean and organize historical patient data for analysis.
- **Action:** Gather data on visit durations, appointment types, demographics. Preprocess to handle missing values and normalize data.

2. Model Training:

- **Reasoning:** Predict appointment durations and patient needs.
- **Action:** Train regression models (e.g., Linear Regression, Random Forest) to forecast visit lengths and classify appointment priorities.

3. Schedule Optimization:

- **Reasoning:** Maximize efficiency and minimize wait times.
- **Action:** Use predictions to create an optimized schedule with buffer times, prioritizing urgent cases using optimization algorithms (e.g., Linear Programming).

