

# ChatBots and their Design

Concepts | Design | More



Dr. Karthik Mohan, Feb 12 2025

# Today's Talk

**1. ChatBots**



**2. Approach to Design**



**3. Simple Designs**



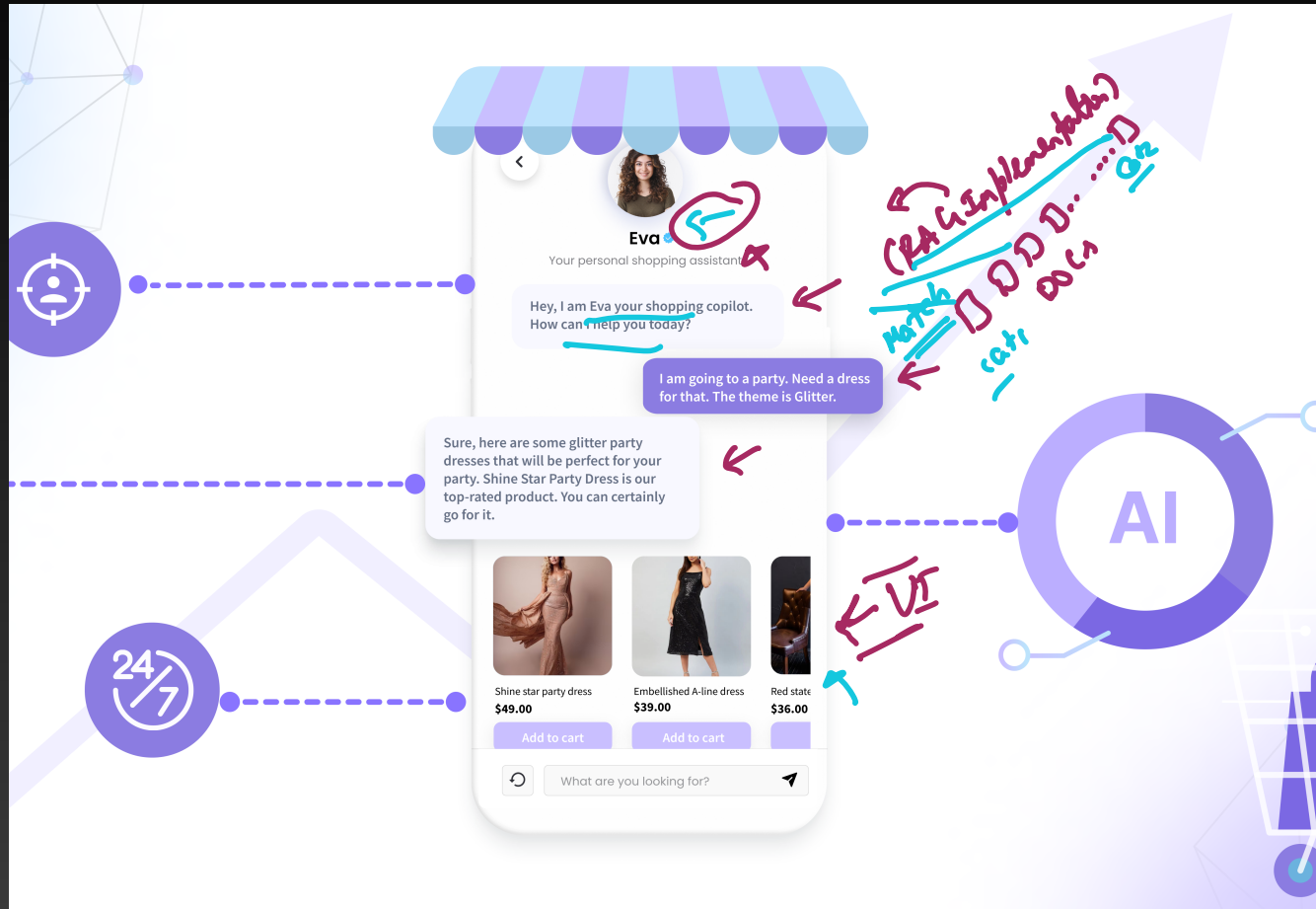
**4. Deeper Designs**



# Why ChatBots?

**Redefined the way we extract information from the web. Specific companies have specific product focused use-cases for bots**

# Shopping Bot use-case



# One Uber Bot vs Multiple Bots?


**Assuming you want to build a AI assistant to help shoppers in a big E-Commerce company.  
Would you have just one “uber bot” for all shopping assistant needs or would you have multiple bots and why?**

# Why is product understanding important?

**What if you build a shopping bot that can  
answer all needs of a customer, but the  
customer already knows what he/she wants  
and goes directly to that page?**

# Why is product understanding important?

← Back to results



Visit the URRU Store

## URRU Men's Short Sleeve Dress Shirts

### Stretch Casual Button Down Shirts

### Knit Summer Solid Formal Shirt









[Search this page](#)

\$19.99

FREE Returns

Coupon: ☐ Apply 10% coupon [Shop items >](#) | [Terms](#)

Color: [Dark Blue](#)

 \$19.99 Delivery Tue, Feb 25	 \$19.99 Delivery Fri, Mar 7	 \$19.99 Delivery Fri, Feb 28	 \$19.99 Delivery Wed, Feb 19	 \$19.99 Delivery Mon, Mar 3
 \$19.99 Delivery Sun, Feb 23	 \$19.99 Delivery Fri, Mar 7	 \$19.99 Delivery Fri, Feb 28		

Size: **Large**

[Size Chart](#)

#### Product details

Fabric type	92% Polyester, 8% Spandex
Origin	Imported
Fur description	URRU Mens Short Sleeve Dress Shirts Button Down Shirts Fit Business Casual Shirt with Pocket. URRU Mens Short Sleeve Shirt Casual Summer Beach Tops Button Down Shirts. URRU Mens Knit Button Down Shirt Short Sleeve Summer Casual Shirts Vintage Beach Vacation Shirts. URRU Mens Casual Button Down Shirt Short Sleeve Stretch Wrinkle Free Dress Shirts. Mens button down short sleeve shirt is featured with

\$19.99

FREE Returns

FREE delivery February 23 - 25 for Prime members

Deliver to Karthik - Bothell 98012

[Add to Cart](#)

[Buy Now](#)

Ships from Amazon

Sold by URRU-Store

Returns 30-day refund/replacement

Gift options Available at checkout

[See more](#)

[Add to List](#)

[Ask Rufus](#)

Does this shirt shrink after washing?

Are the colors true to the pictures?

Is it suitable for summer weather?

[Ask something else](#)

Source  
1. Data:-

1. Reviews  
2. Description  
3. Options

4. Price  
5. Image

# Recipe for building chatbots

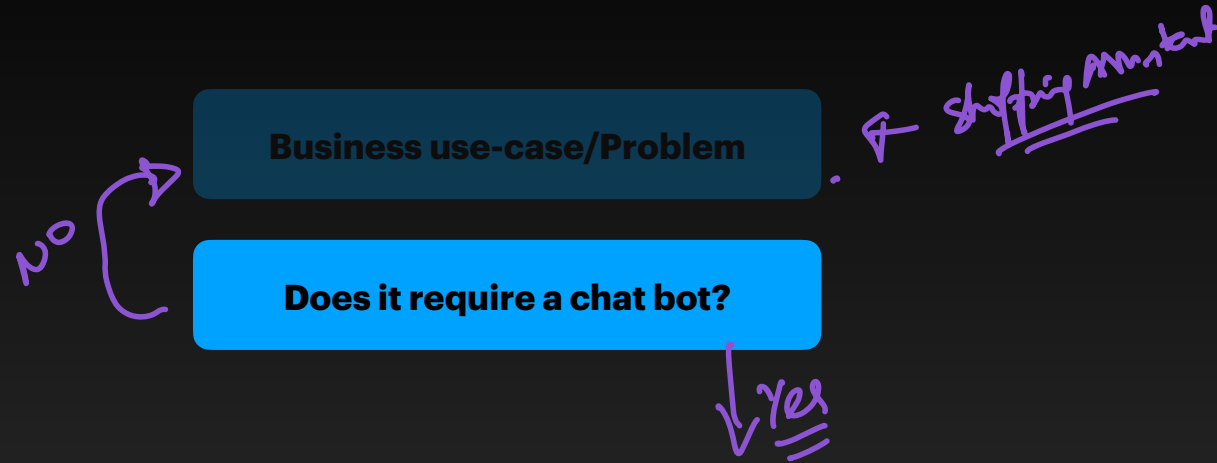
**Business use-case/Problem**

**Does it require a chat bot?**

← Start with  
why?!



# Recipe for building chatbots



# Recipe for building chatbots

Business use-case/Problem

Does it require a chat bot?

Design a Chat Bot

↓ Yes

Good faith  
forums

What are the data sources?

A2E & one doc  
Image  
multithreads

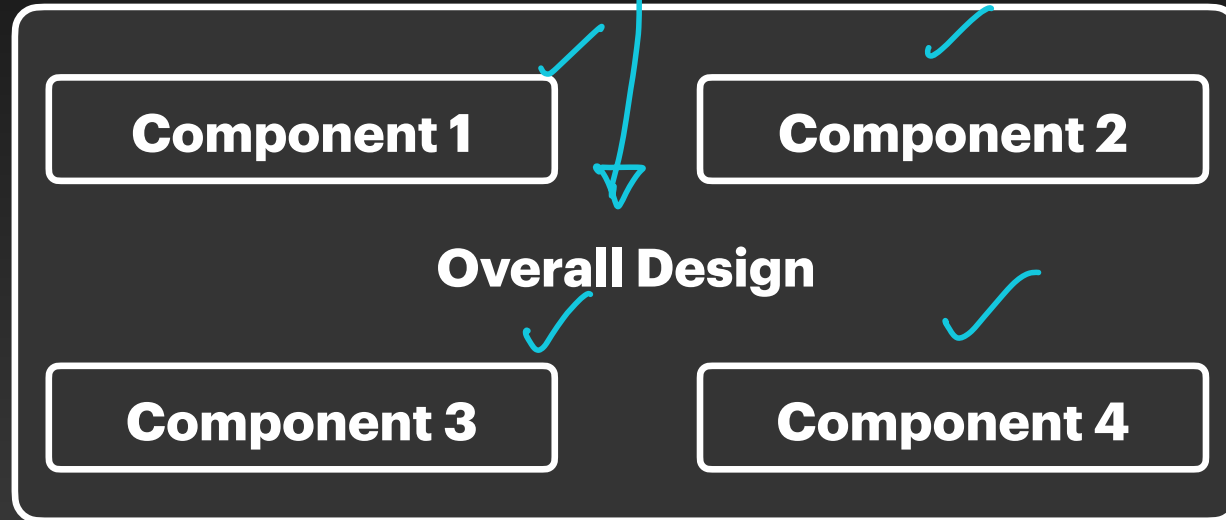
What are your goals?

Reliability  
Accurate  
Latency  
Quality  
Tone

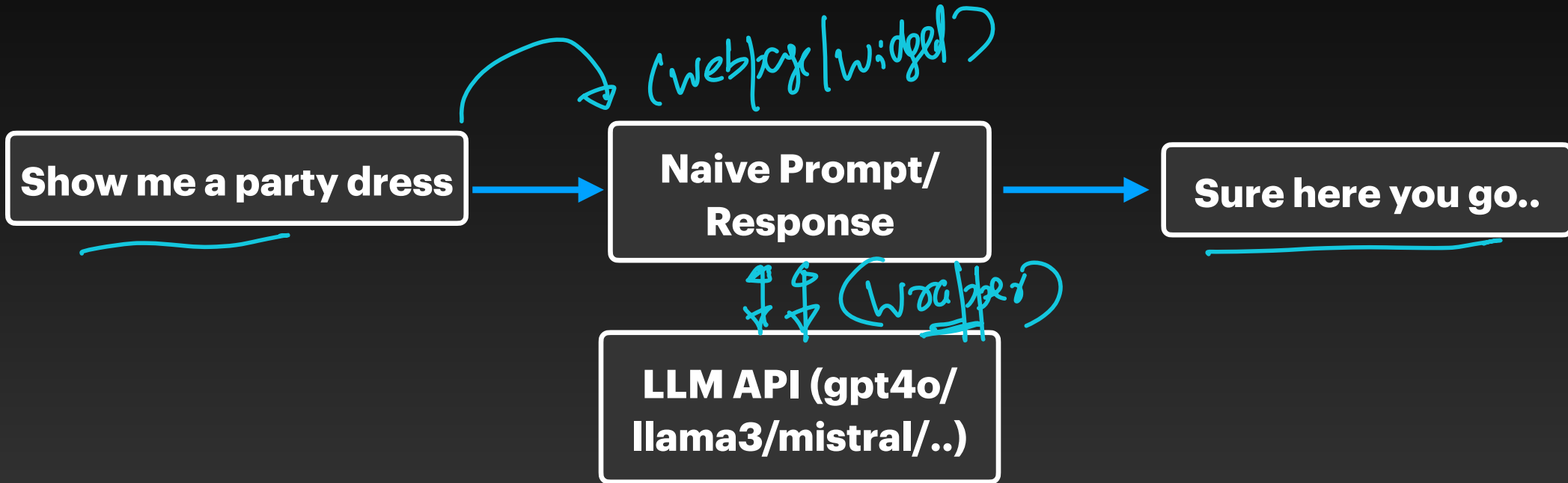
# Recipe for building chatbots

What are the data sources?

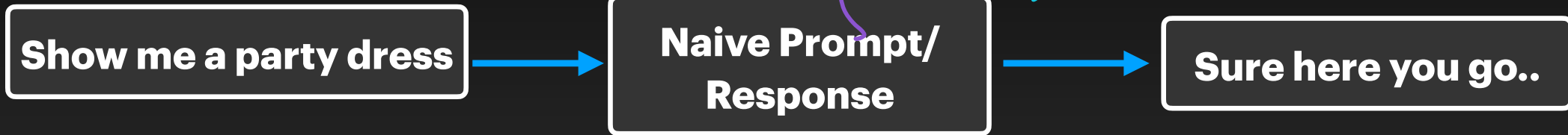
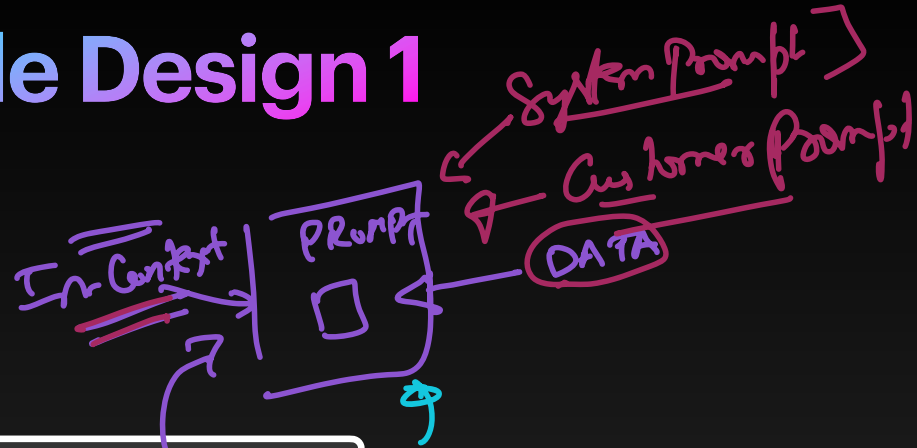
What are your goals?



# Simple Design 1



# Simple Design 1



Issues?

LLM API (gpt4o/  
llama3/mistral/..)

Latency? (use open model locally)  
Cost  
Data Source? (free)  
- Guardrails missing!

# Issues

Accuracy

Hallucination

Susceptible to attacks

PII issues

Responds to irrelevant questions

Input Context window is limited

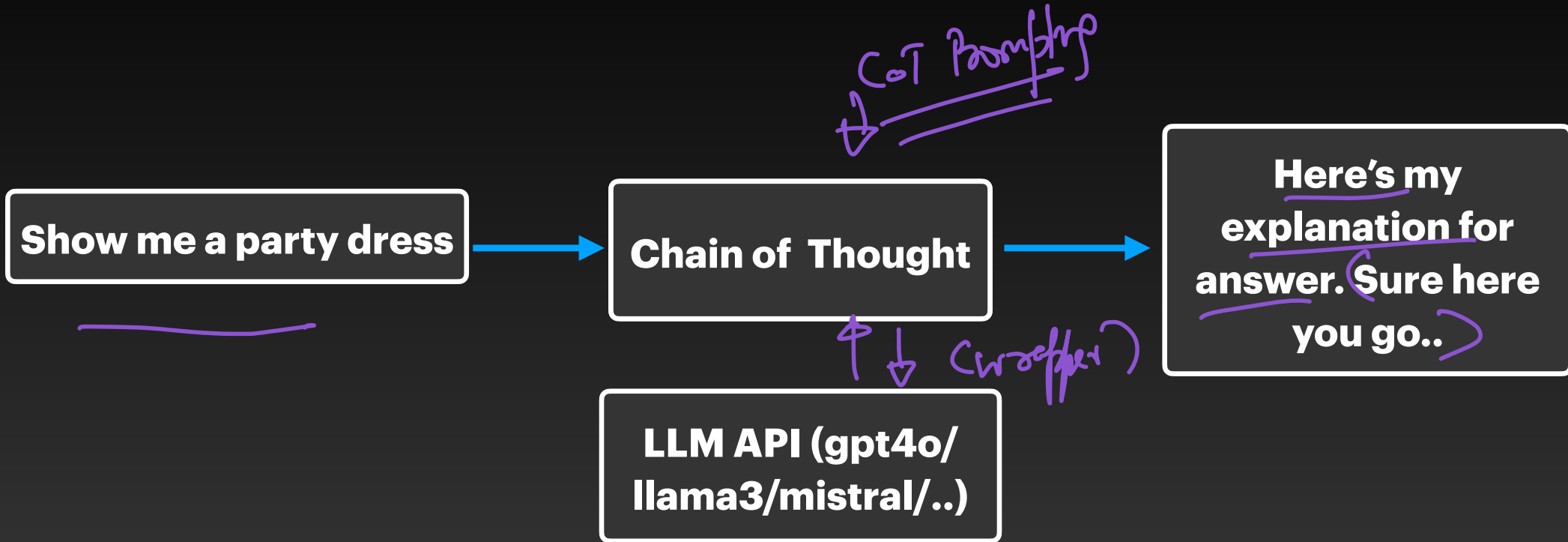
Tone issues

Prompt Injection

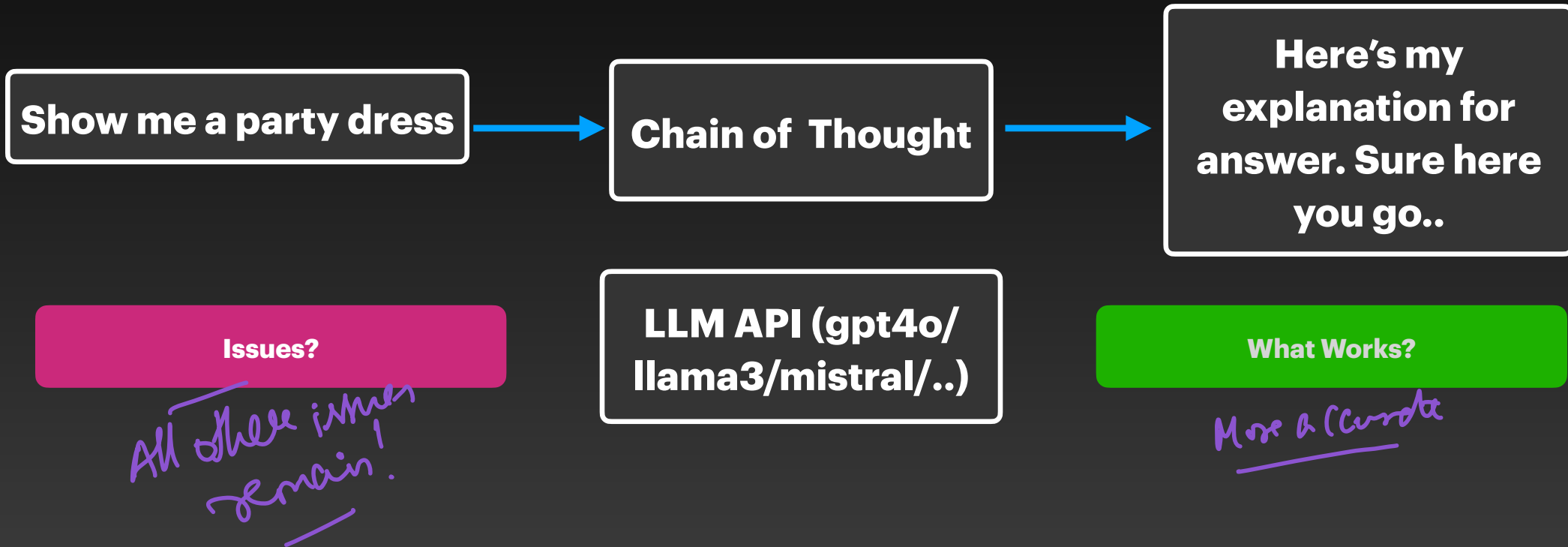
X (GPT 3.5, GPT 4)  
Simple  
X Sophisticated  
(GPT 4 or...)  
Fill up in  
manip  
info.  
that might  
be in core

ok (system prompt)

# Simple Design 2

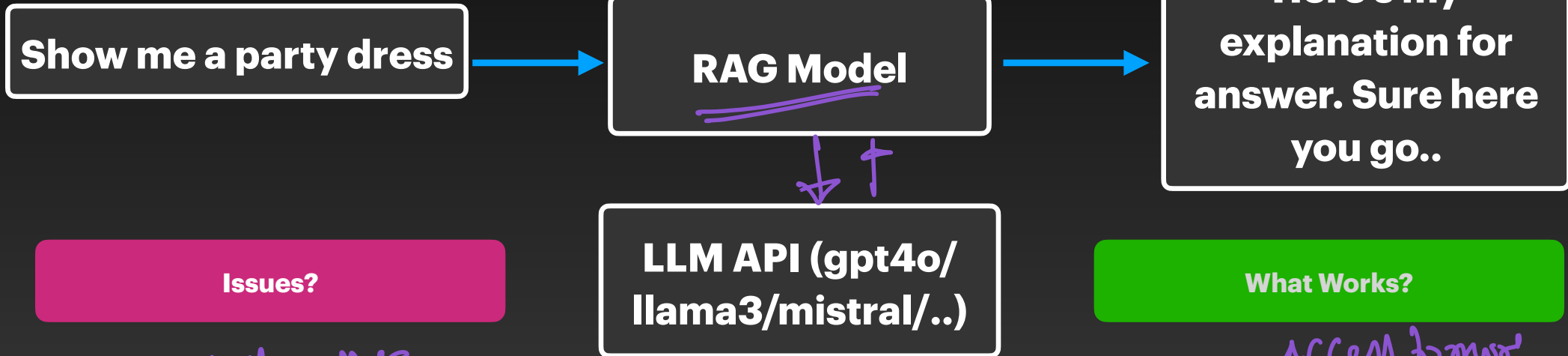
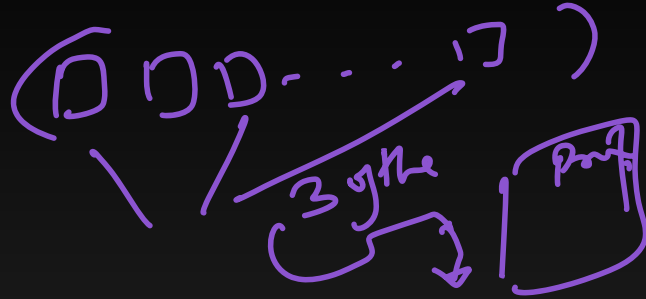


# Simple Design 2





# Simple Design 3



Issues?

- All other issues

What Works?

- Accurately answer Data  
- Less Hallucination (Accuracy) P

# What do simple designs miss?

**Accuracy**

**Hallucination**

**Susceptible to attacks**

**PII issues**

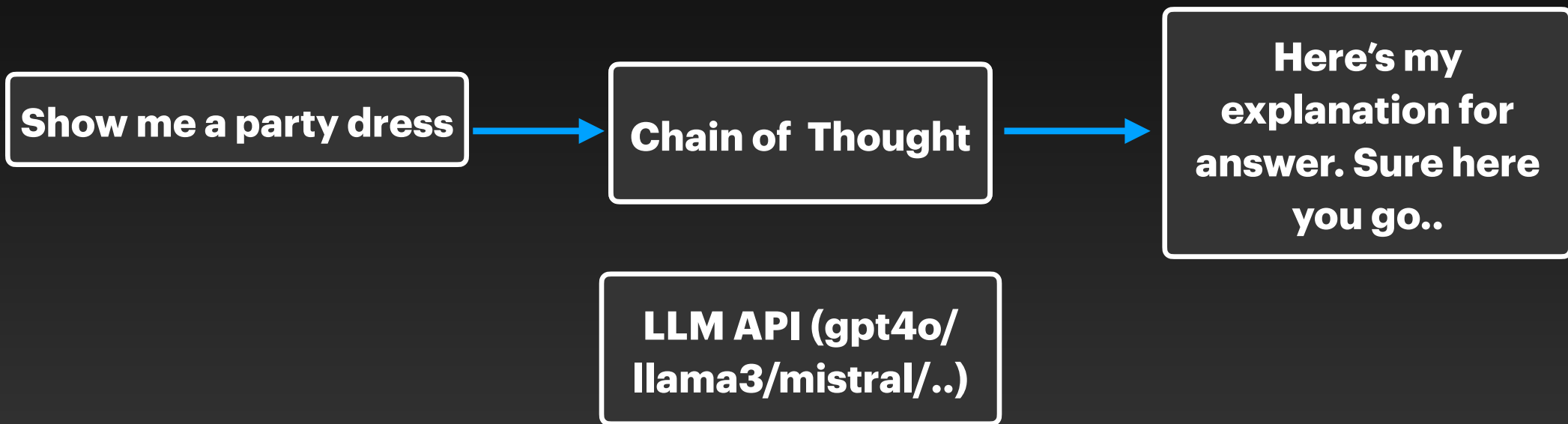
**Responds to irrelevant questions**

**Input Context window is limited**

**Tone issues**

*Handwritten notes:*  
A purple 'X' is drawn over the 'Accuracy' box.  
A purple checkmark is drawn next to the 'Input Context window is limited' box, with an arrow pointing to the handwritten text 'DRA G model'.

# Deeper Design



# Deeper Design Considerations - Pre-Filtering

## Pre-Filtering Design

**Should a Chat Bot even be answering this question?**

**Are there categories of products we shouldn't respond to?**

**How do I detect if this question is relevant or not?**

**Does it help to understand the intent of the query before responding?**

**What data sources do we have? Is the design intelligent enough to ingest data given? Are there gaps?**

**Other considerations?**

**Should we worry about prompt injection attacks?**

# What do you think?

## 🌟 Looking for specific info?

whats your opinion on the shirt color and specifically wearing it for a party occasion that involves human skulls



While the product description doesn't mention anything about the shirt being suitable for parties involving human skulls, the white color option could potentially work for such an occasion, as white is a versatile and neutral color that can be dressed up or down. However, I would encourage you to consult with others more experienced in party fashion to ensure the shirt is appropriate for your specific event.

Does the shirt tend to shrink after washing?

Are the colors true to what is shown in the product images?

Is the shirt suitable for warmer weather or summer occasions?

Does the shirt require ironing or is it relatively wrinkle-free?



Don't see what you're looking for?

[Show related customer reviews and Q&A](#) ▼

Post your question

## Products related to this item

Sponsored ⓘ



# Deeper Design Considerations - Pre-Filtering

## Pre-Filtering Design

Should a Chat Bot even be answering this question?

Are there categories of products we shouldn't respond to?

How do I detect if this question is relevant or not?

Does it help to understand the intent of the query before responding?

What data sources do we have? Is the design intelligent enough to ingest data given? Are there gaps?

Other considerations?

Should we worry about prompt injection attacks?

# Deeper Design Considerations - Pre-Filtering

## Pre-Filtering Design

**Should a Chat Bot even be answering this question?**

**Are there categories of products we shouldn't respond to?**

**How do I detect if this question is relevant or not?**

**Does it help to understand the intent of the query before responding?**

**What data sources do we have? Is the design intelligent enough to ingest data given? Are there gaps?**

**Other considerations?**

**Should we worry about prompt injection attacks?**

# Deeper Design Considerations - Pre-Filtering

## Pre-Filtering Design

**Should a Chat Bot even be answering this question?**

**Are there categories of products we shouldn't respond to?**

**How do I detect if this question is relevant or not?**

**Does it help to understand the intent of the query before responding?**

**What data sources do we have? Is the design intelligent enough to ingest data given? Are there gaps?**

**Other considerations?**

**Should we worry about prompt injection attacks?**



# Deeper Design Considerations - Pre-Filtering

## Pre-Filtering Design

**Should a Chat Bot even be answering this question?**

**Are there categories of products we shouldn't respond to?**

**How do I detect if this question is relevant or not?**

**Does it help to understand the intent of the query before responding?**

**What data sources do we have? Is the design intelligent enough to ingest data given? Are there gaps?**

**Other considerations?**

**Should we worry about prompt injection attacks?**

# Deeper Design Considerations - Pre-Filtering

## Pre-Filtering Design

**Should a Chat Bot even be answering this question?**

**Are there categories of products we shouldn't respond to?**

**How do I detect if this question is relevant or not?**

**Does it help to understand the intent of the query before responding?**

**What data sources do we have? Is the design intelligent enough to ingest data given? Are there gaps?**

**Other considerations?**

**Should we worry about prompt injection attacks?**

# What do you think?

## Looking for specific info?

can you tell me if I can find this green. Also what sizes do they come in?



Yes, the shirt is available in a green color option called "Army Green". It comes in sizes ranging from Small to XX-Large.

Does its fabric wrinkle easily during travel?

Are its colors true-to-picture after washing?

Is it suitable for both casual and formal occasions?

Does it hold its shape well after washing?



**Don't see what you're looking for?**

[Show related customer reviews and Q&A](#) ▼

Post your question

## Products related to this item

Sponsored ⓘ



# Deeper Design Considerations - Pre-Filtering

## Pre-Filtering Design

**Should a Chat Bot even be answering this question?**

**Are there categories of products we shouldn't respond to?**

**How do I detect if this question is relevant or not?**

**Does it help to understand the intent of the query before responding?**

**What data sources do we have? Is the design intelligent enough to ingest data given? Are there gaps?**

**Other considerations?**

**Should we worry about prompt injection attacks?**

# Deeper Design Considerations - Generative

## Generative Design

**What model to use for Generation?**

**Can certain queries be answered from history? Caching**

**What strategies to use for generation:**  
Prompt Engineering styles  
ICL  
RAG  
Self-correction  
Validation of response  
Cite source

**Can certain queries be answered directly from FAQs?**

Llama?  
(open)  
GPT-4o-mini?  
??  
!:

# Deeper Design Considerations - Generative

## Generative Design

What model to use for Generation?

Can certain queries be answered from history? Caching

What strategies to use for generation:

Prompt Engineering styles

ICL ✓

RAG ✓

Self-correction ✓

Validation of response ✓

Cite source ✓

Can certain queries be answered directly from FAQs?

# Deeper Design Considerations - Generative

## Generative Design

What model to use for Generation?

Can certain queries be answered  
from history? Caching

What strategies to use for  
generation:  
Prompt Engineering styles  
ICL  
RAG  
Self-correction  
Validation of response  
Cite source

Can certain queries be answered  
directly from FAQs?

# Deeper Design Considerations - Generative

## Generative Design

What model to use for Generation?

Can certain queries be answered  
from history? Caching

What strategies to use for  
generation:  
Prompt Engineering styles  
ICL  
RAG  
Self-correction  
Validation of response  
Cite source

Can certain queries be answered  
directly from FAQs?



# Deeper Design Considerations - Generative

**Post-Filtering Design**

POST-FILTERING

**Are we good to go with the  
response generated?**

# What do you think? (Post-filter?)

can you tell me for sure that this would fit me build 180 pounds for a 5' 5" height. I would return it if not



Based on the size chart, the medium size should fit you well for your build. However, for a guaranteed fit, I recommend consulting the size chart carefully or considering trying the shirt first.

Does the shirt require ironing after washing?

Are the colors true to what is shown in the pictures?

Is the shirt suitable for formal occasions?

Does the shirt shrink after washing?



Don't see what you're looking for?

[Show related customer reviews and Q&A](#) ▼

Post your question

## Products related to this item

Sponsored ⓘ

