

PENNING POWERFUL PROMPTS WITH EXAMPLES IN SPRING AI

Craig Wals

 @habuma

 @habuma

 @habuma.bsky.social

 @habuma@fosstodon.org

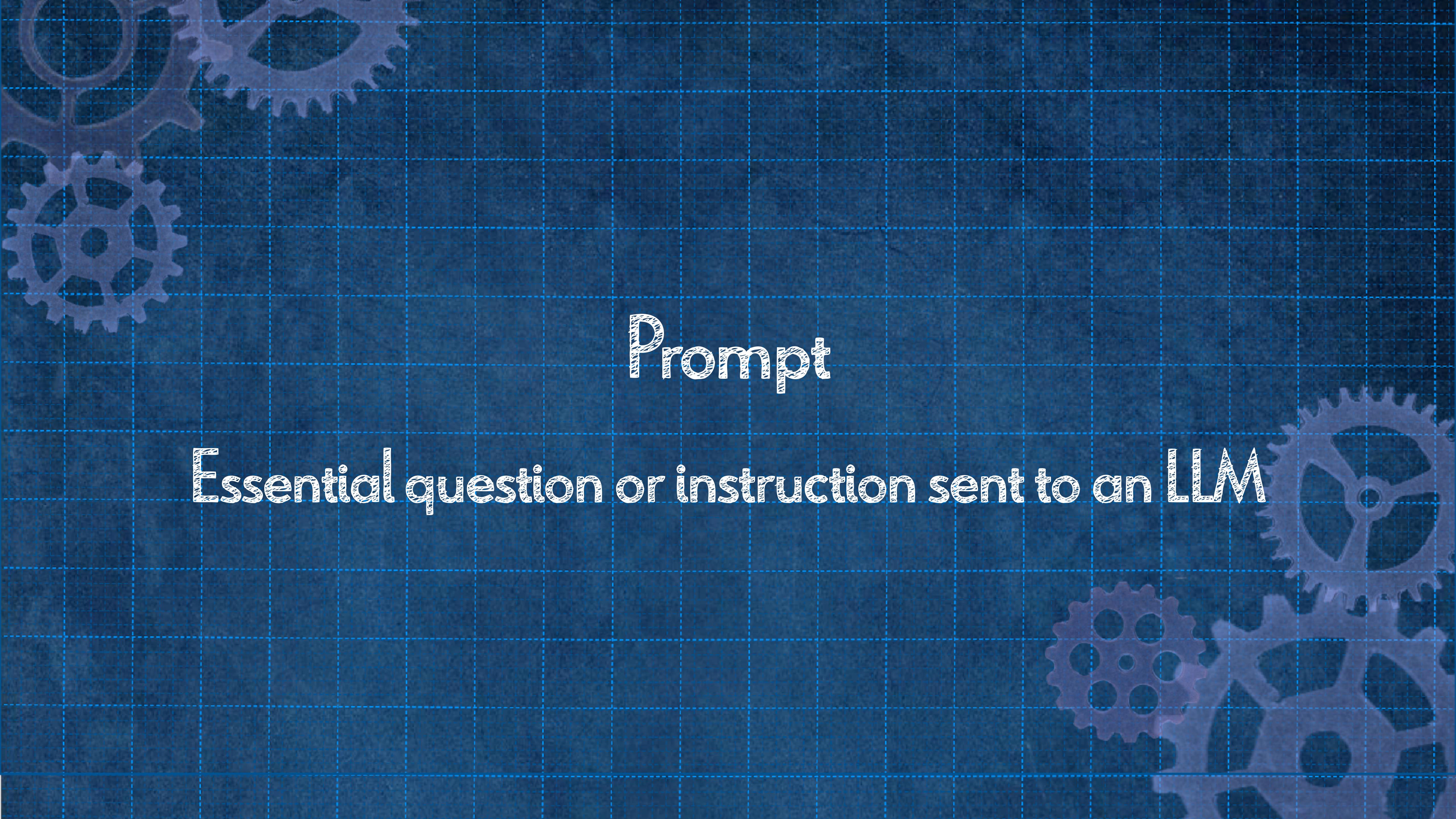
<https://www.habuma.com>

 craig@habuma.com



What we
have time to
cover

What we could talk about...



Prompt

Essential question or instruction sent to an LLM

Anatomy of a Prompt (OpenAI)

Model Selection



```
{  
  "model": "gpt-4.1-nano",  
  "messages": [  
    {  
      "content": "Talk like a pirate",  
      "role": "system"  
    },  
    {  
      "content": "Tell me a joke",  
      "role": "user"  
    }  
  "temperature": 0.7,  
  "stream": false,  
  "tools": [  
    {  
      "type": "function",  
      "function": {  
        ...  
      }  
    }  
  ]  
}
```

Options

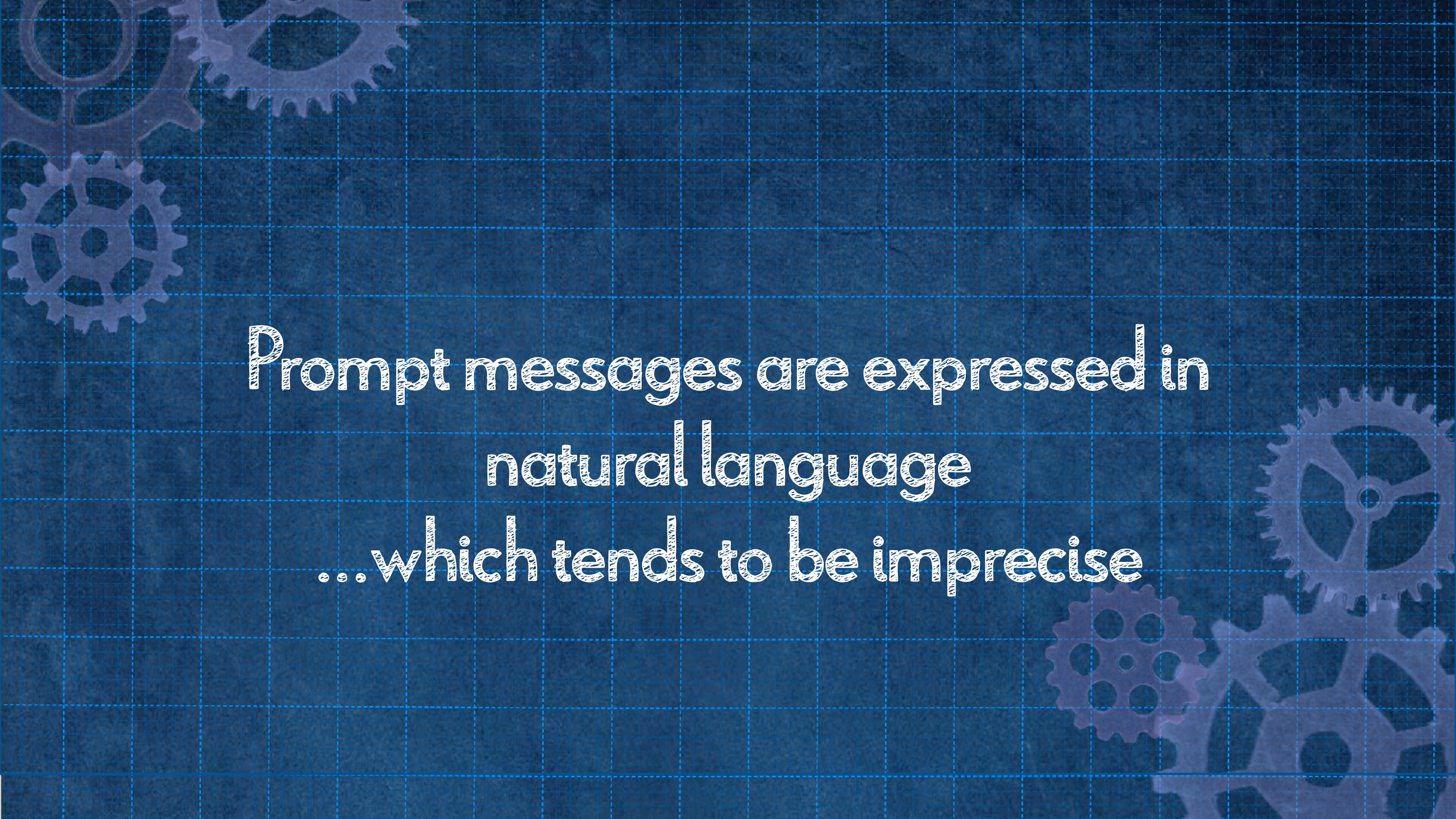


Messages

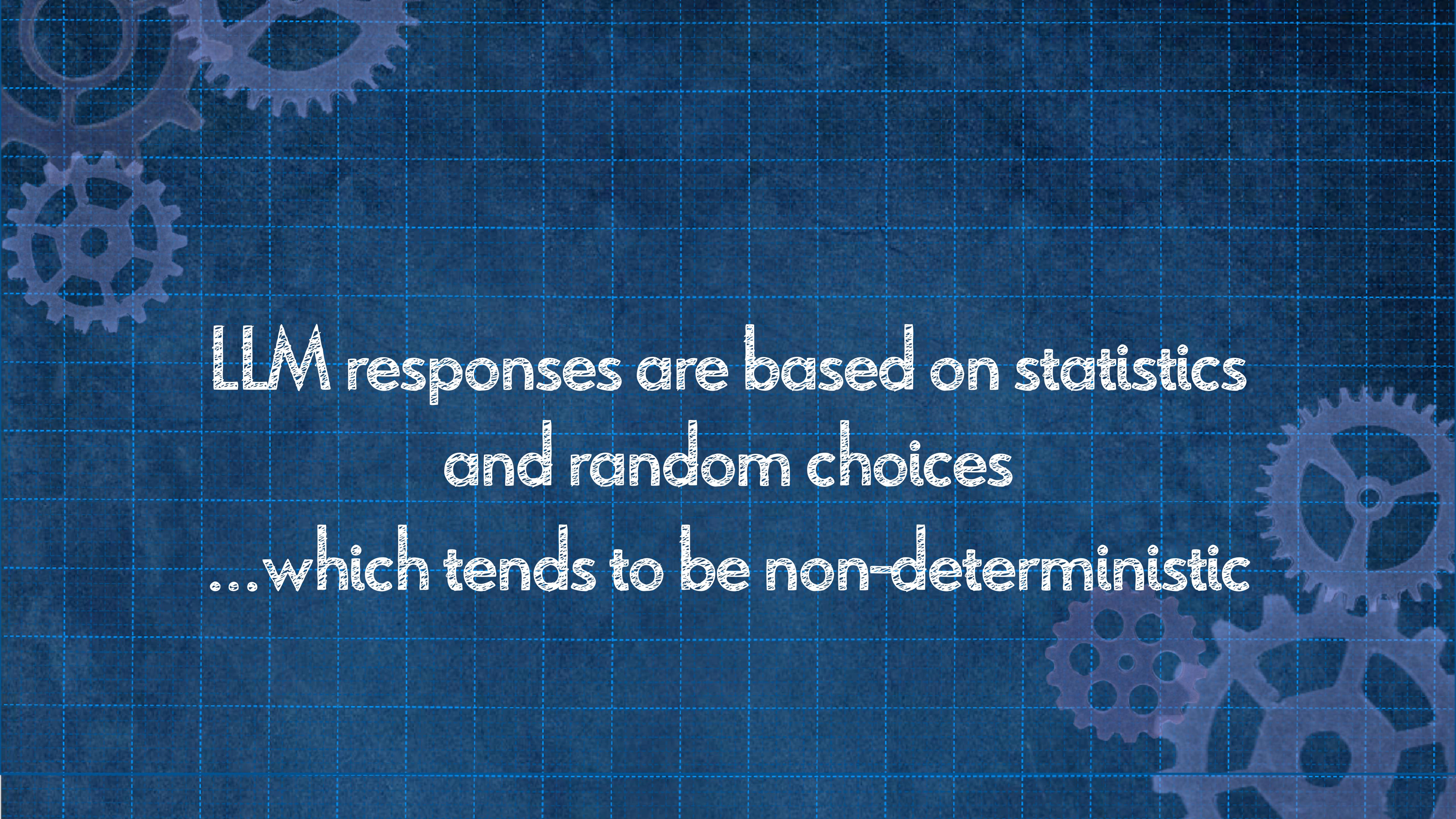


Tools
(Optional)

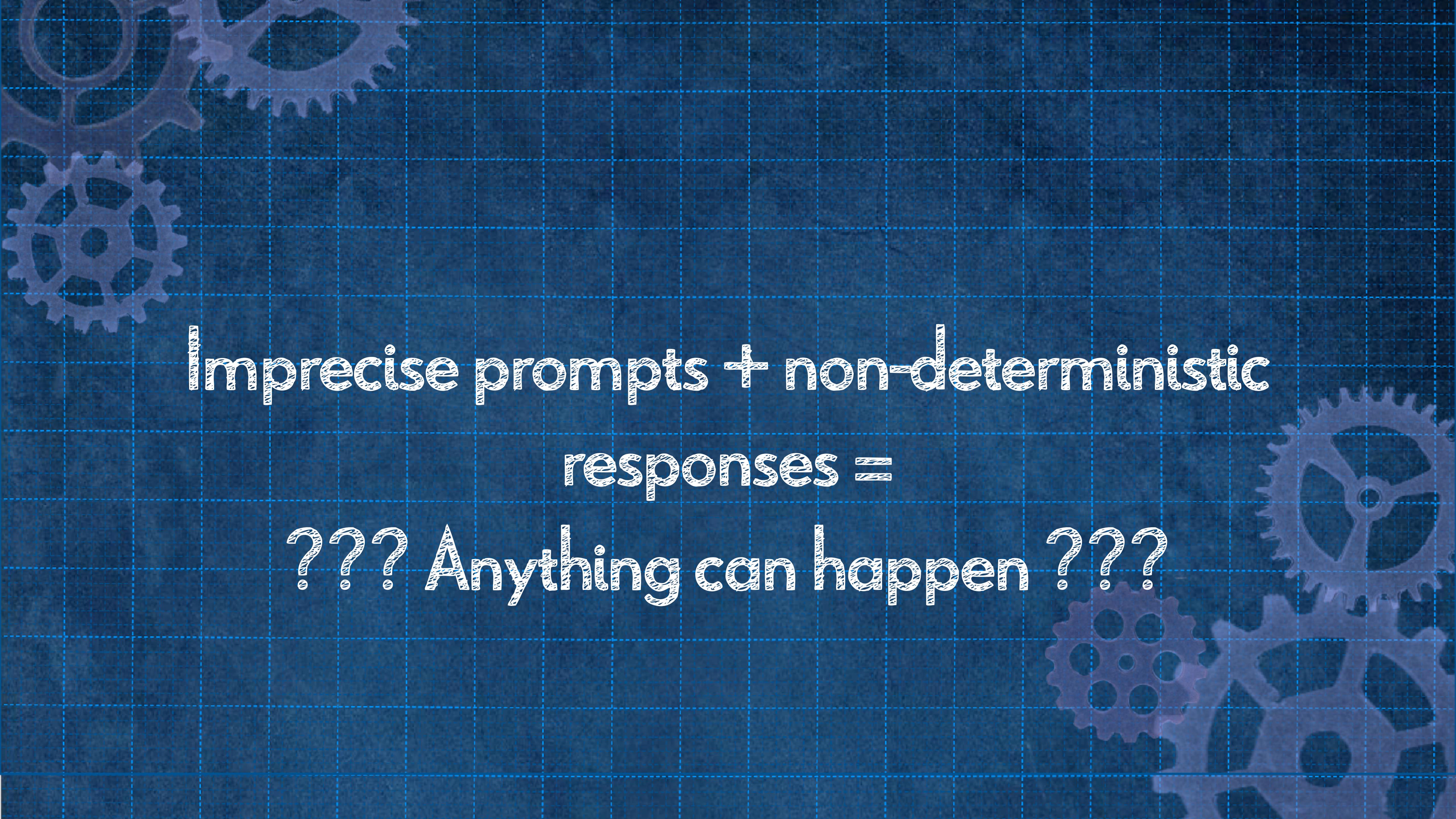




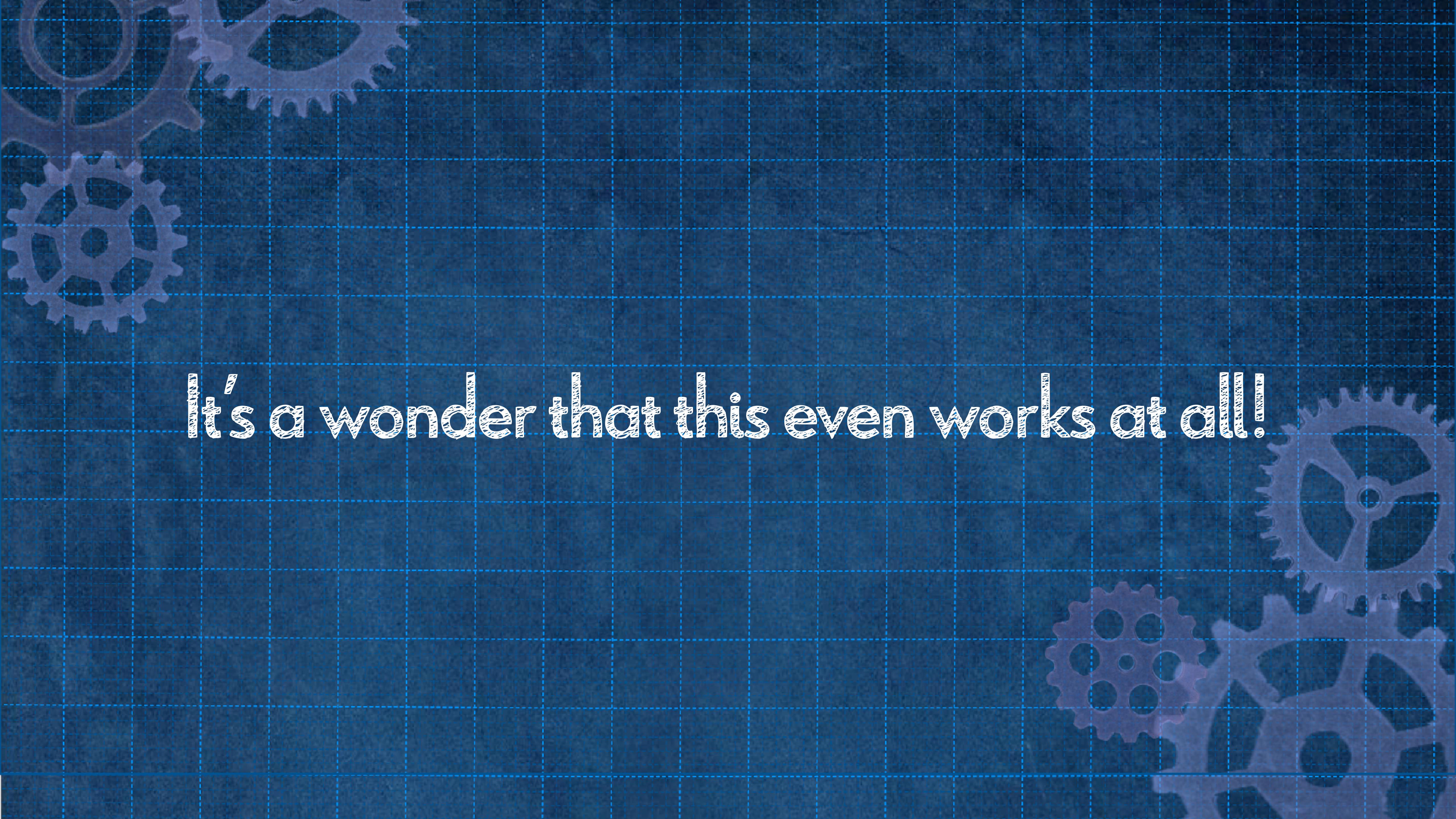
Prompt messages are expressed in
natural language
...which tends to be imprecise



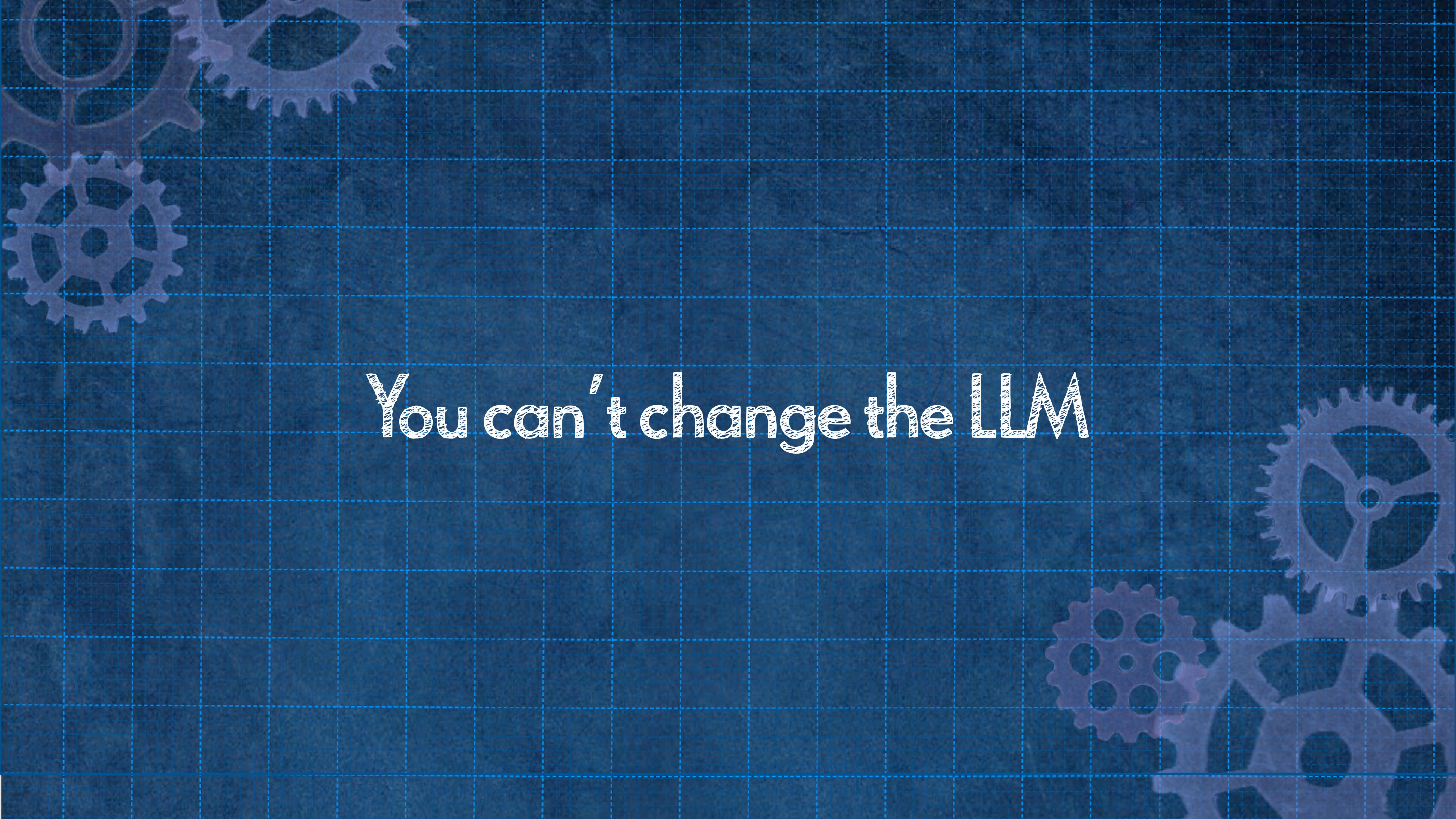
LLM responses are based on statistics
and random choices
... which tends to be non-deterministic



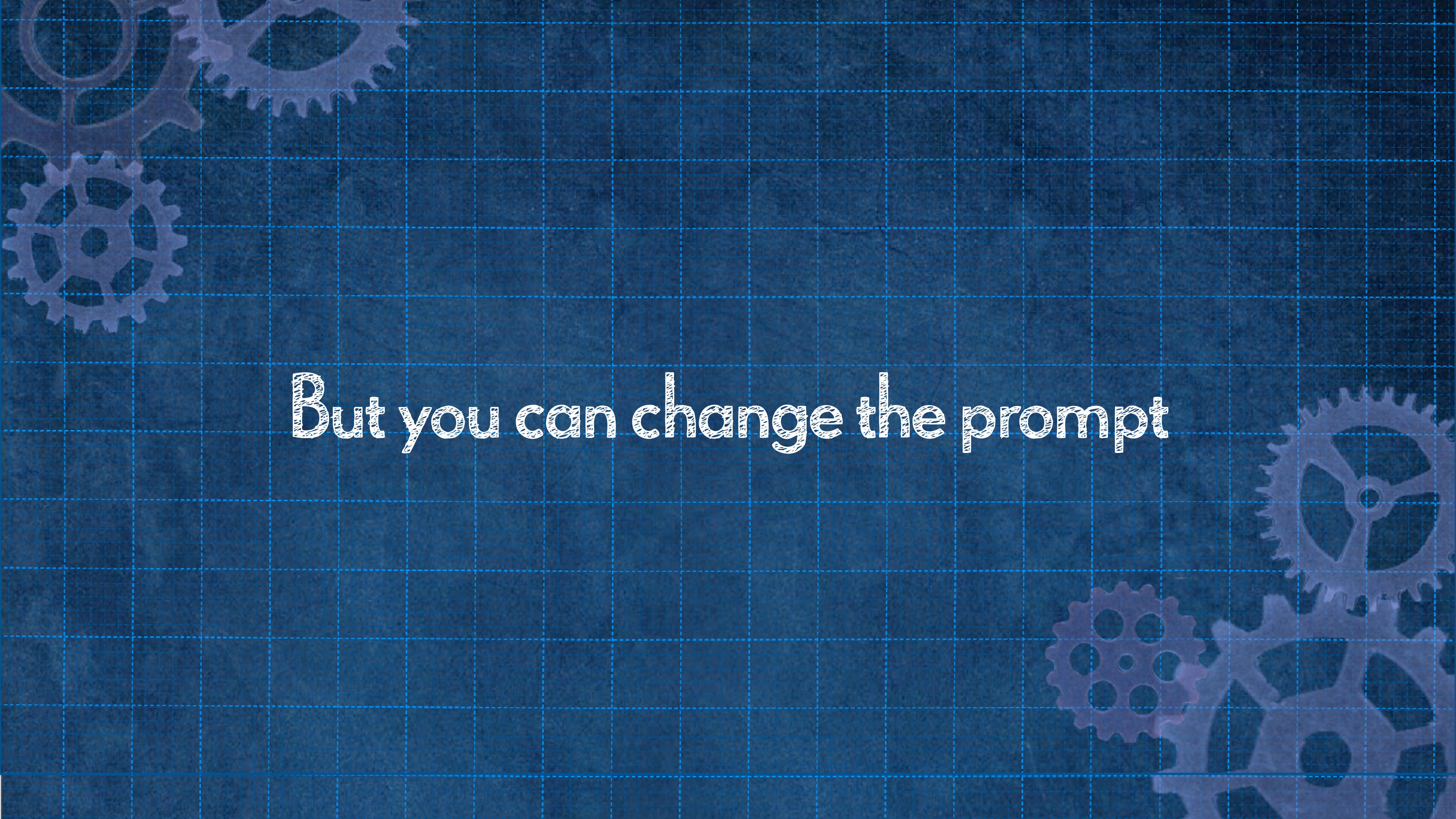
Imprecise prompts + non-deterministic
responses =
??? Anything can happen ???



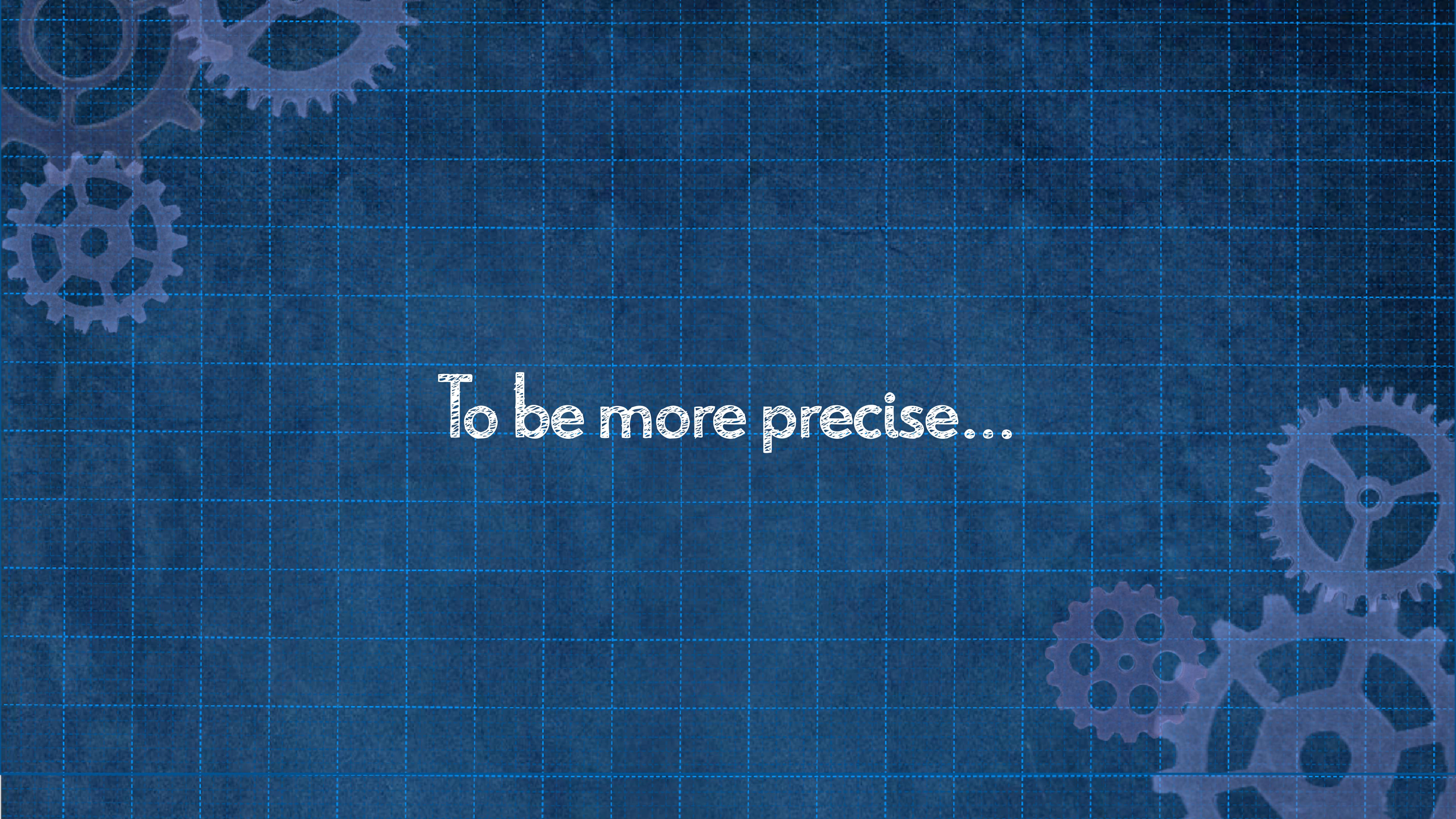
It's a wonder that this even works at all!



You can't change the LLM

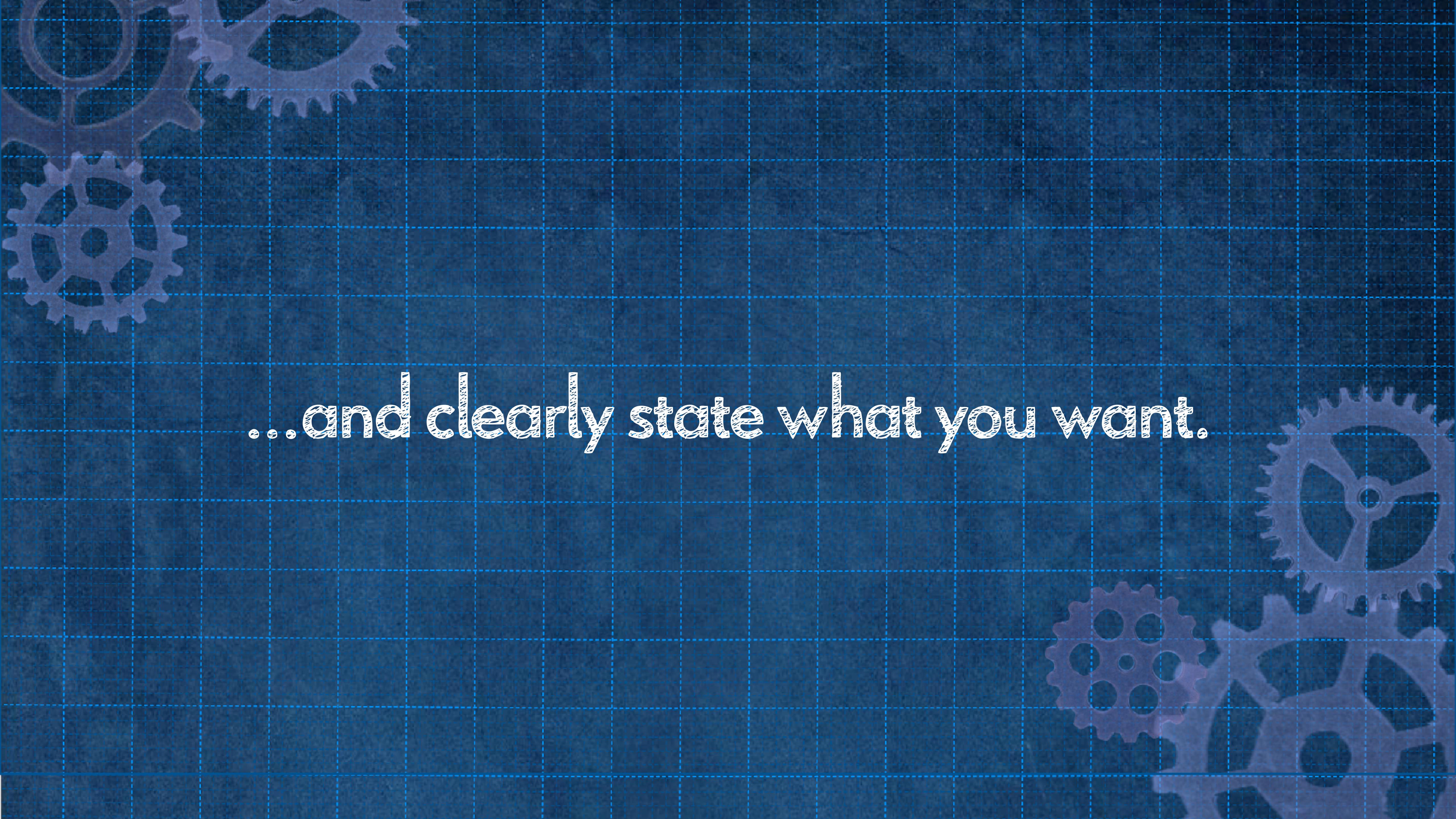


But you can change the prompt

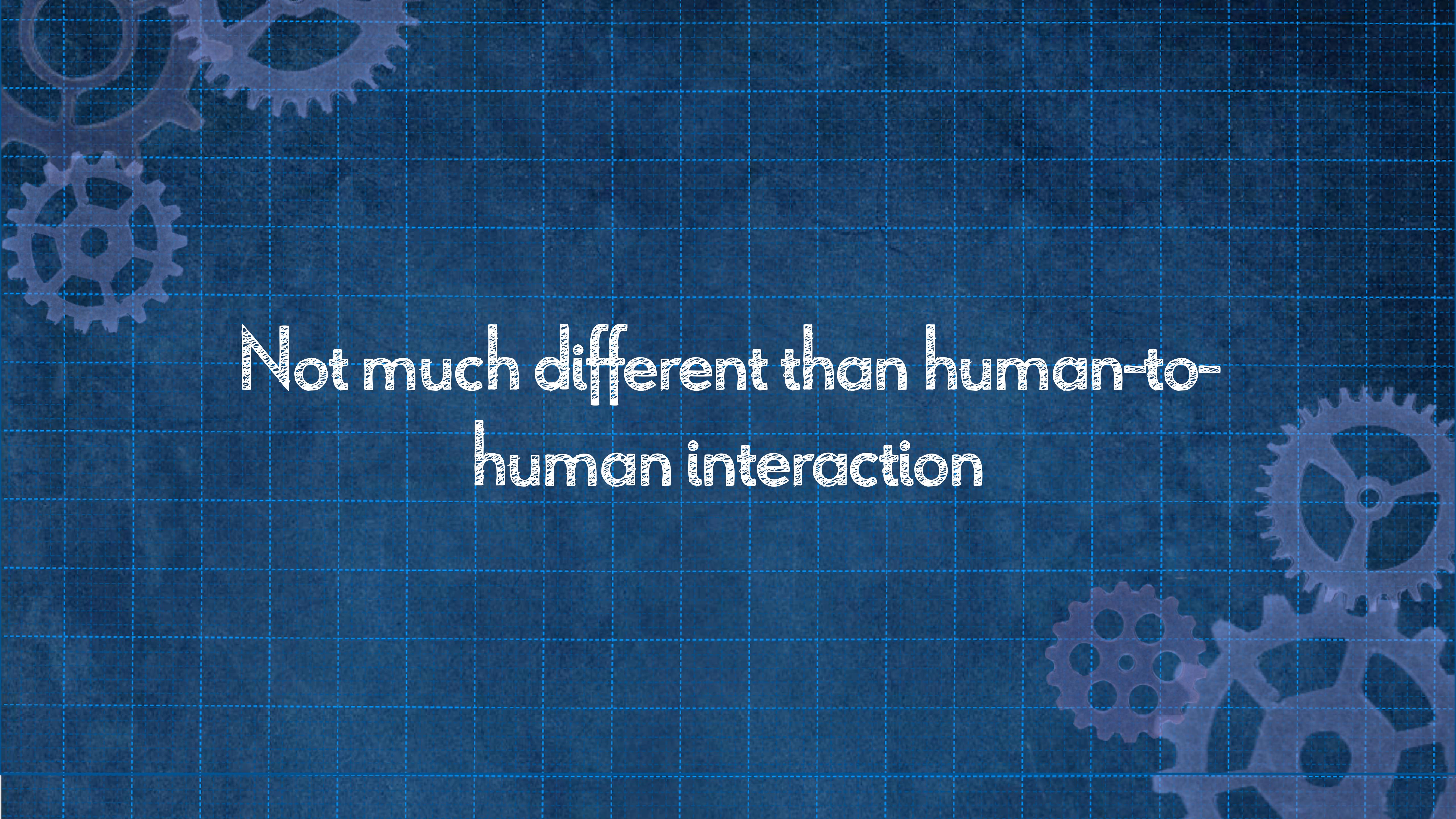


To be more precise...

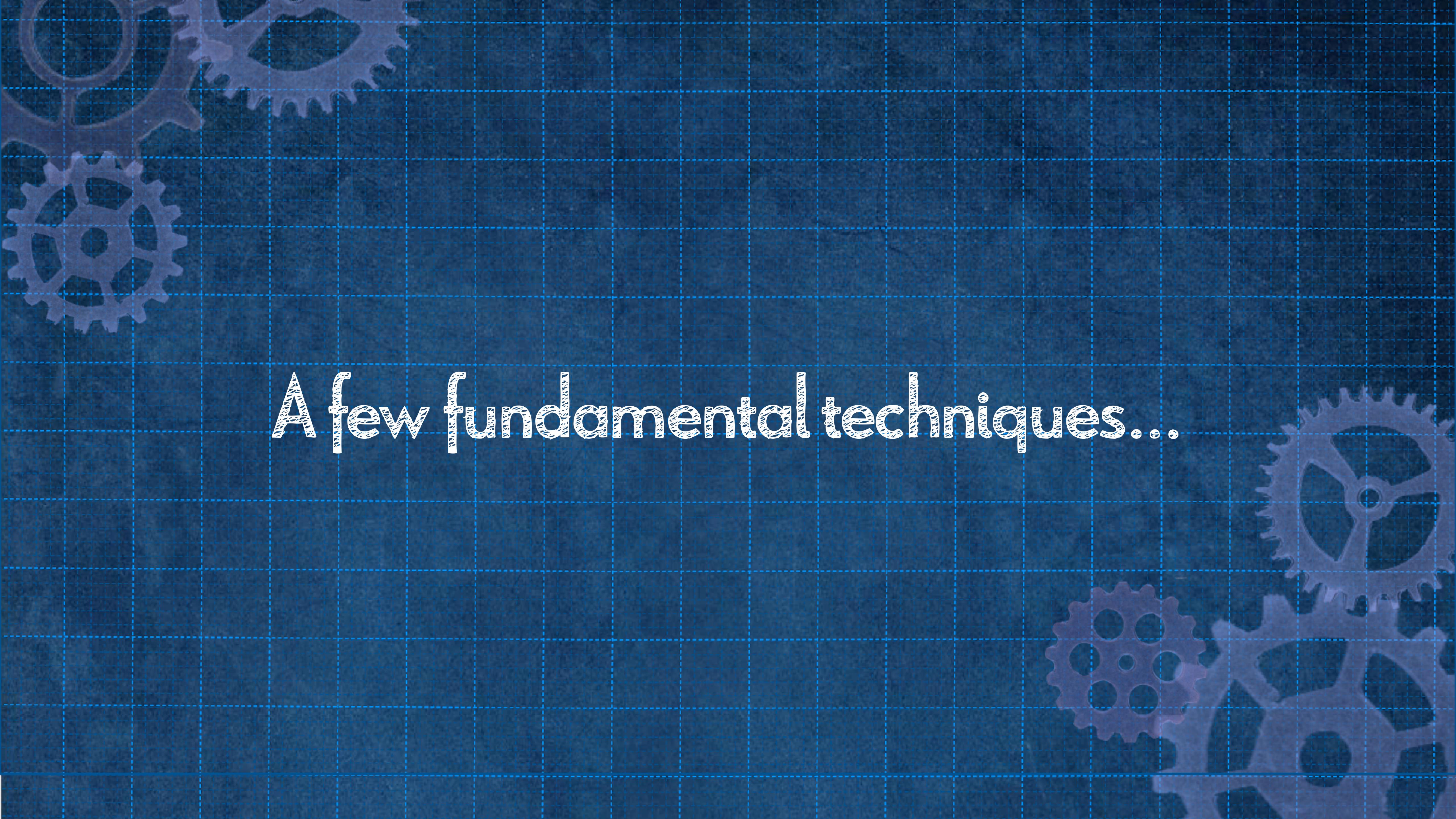
...specific...



...and clearly state what you want.

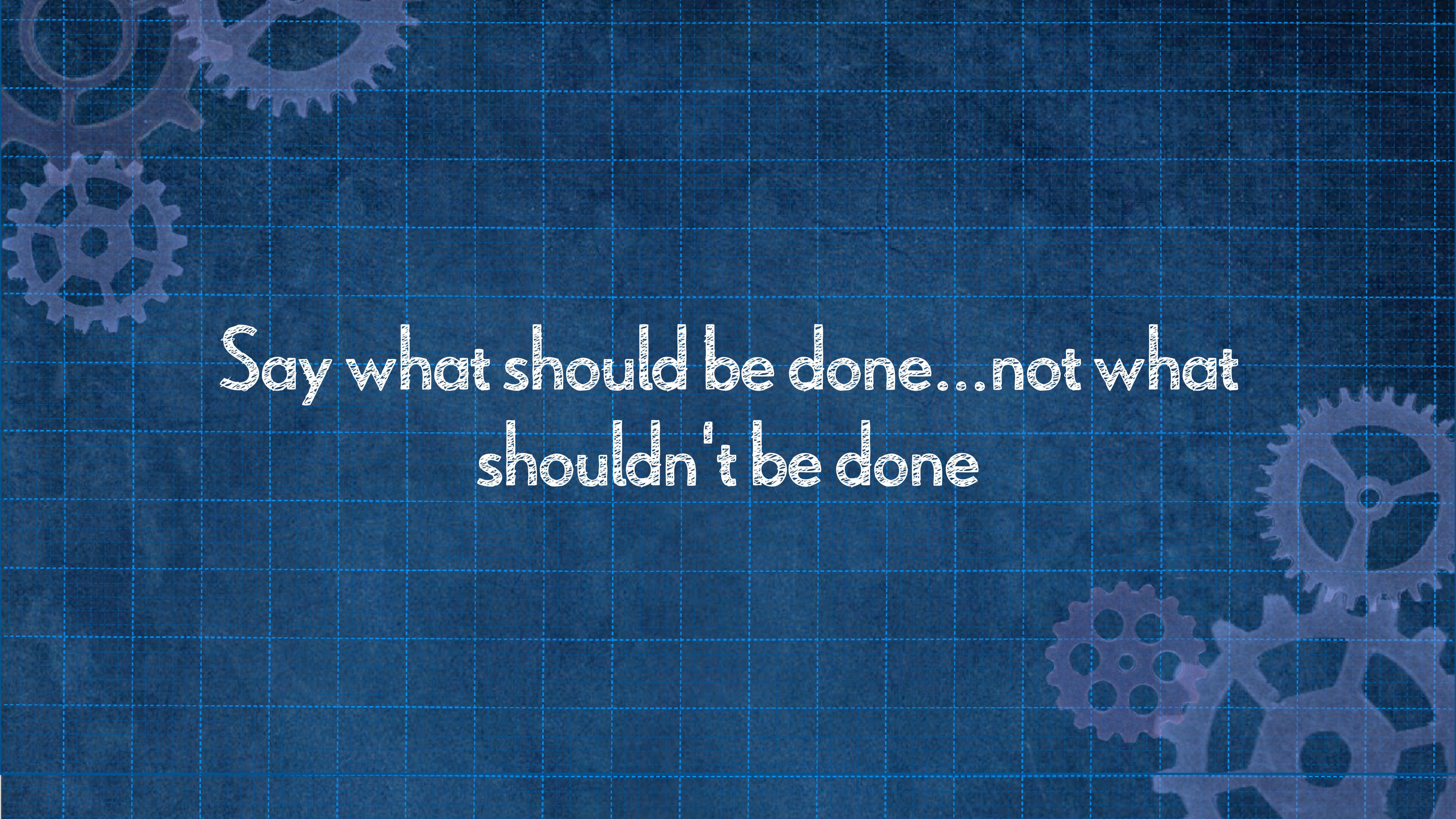


Not much different than human-to-human interaction

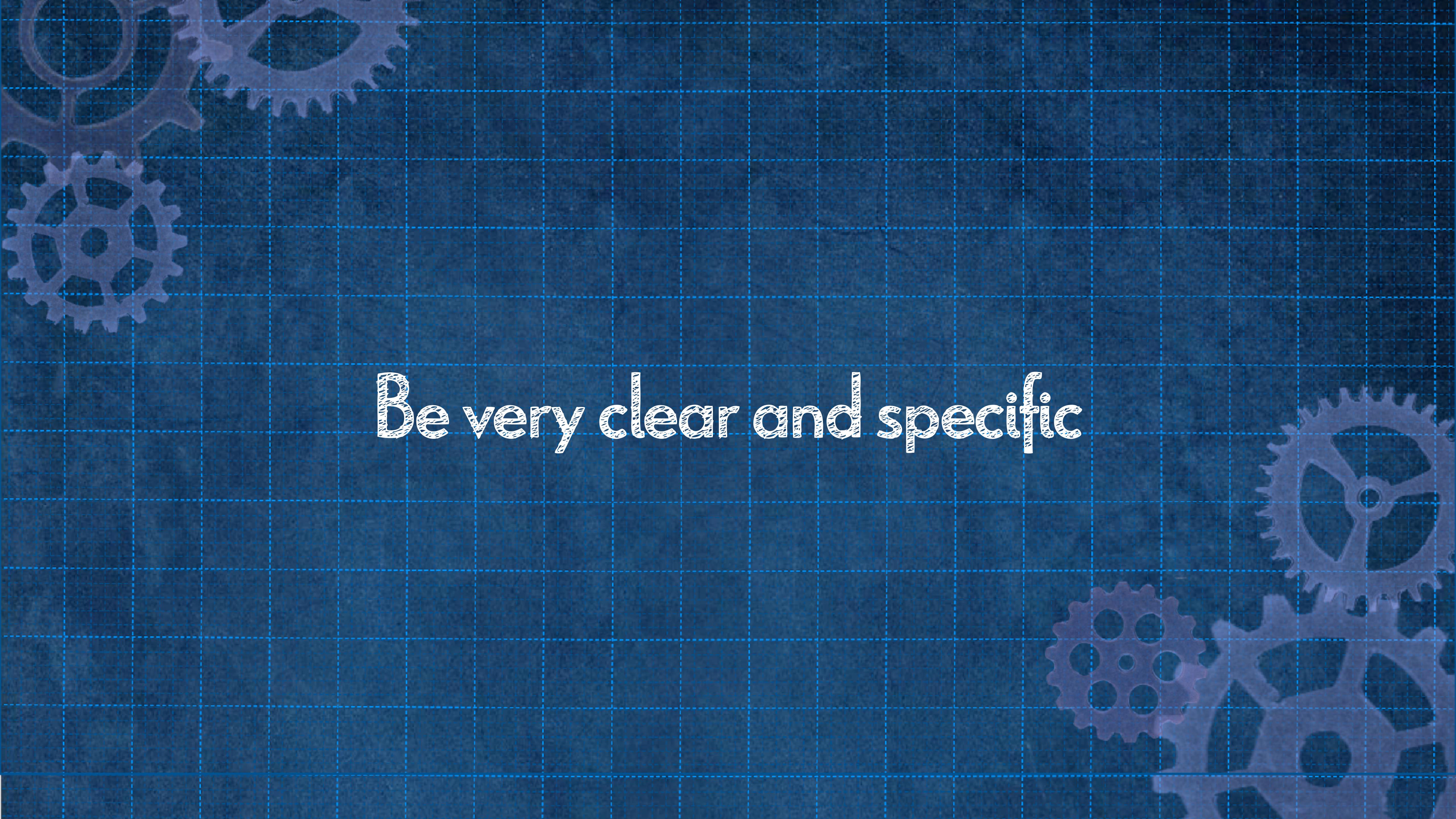


A few fundamental techniques...

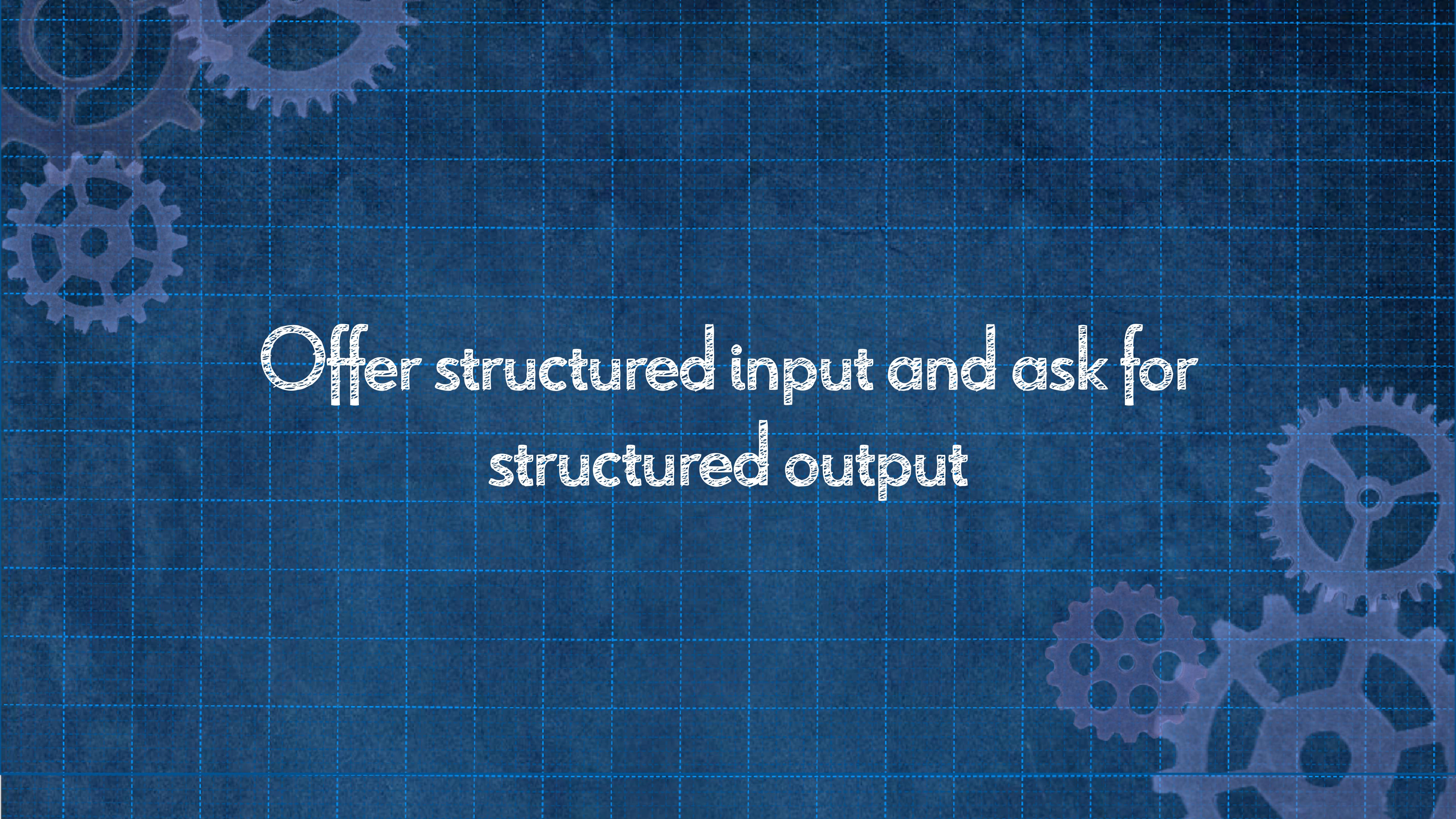
(Disclaimer: YMMV)



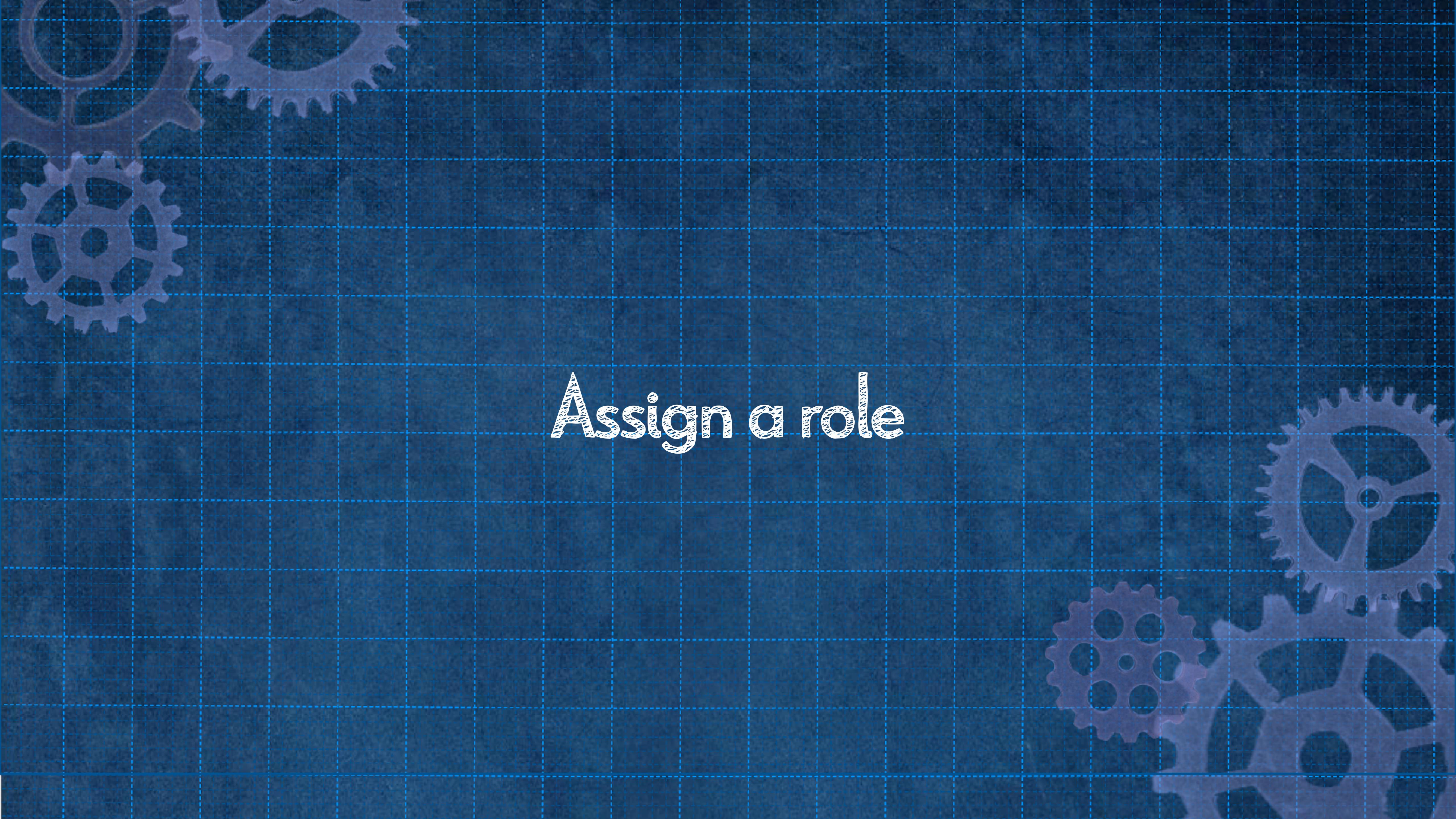
Say what should be done...not what
shouldn't be done



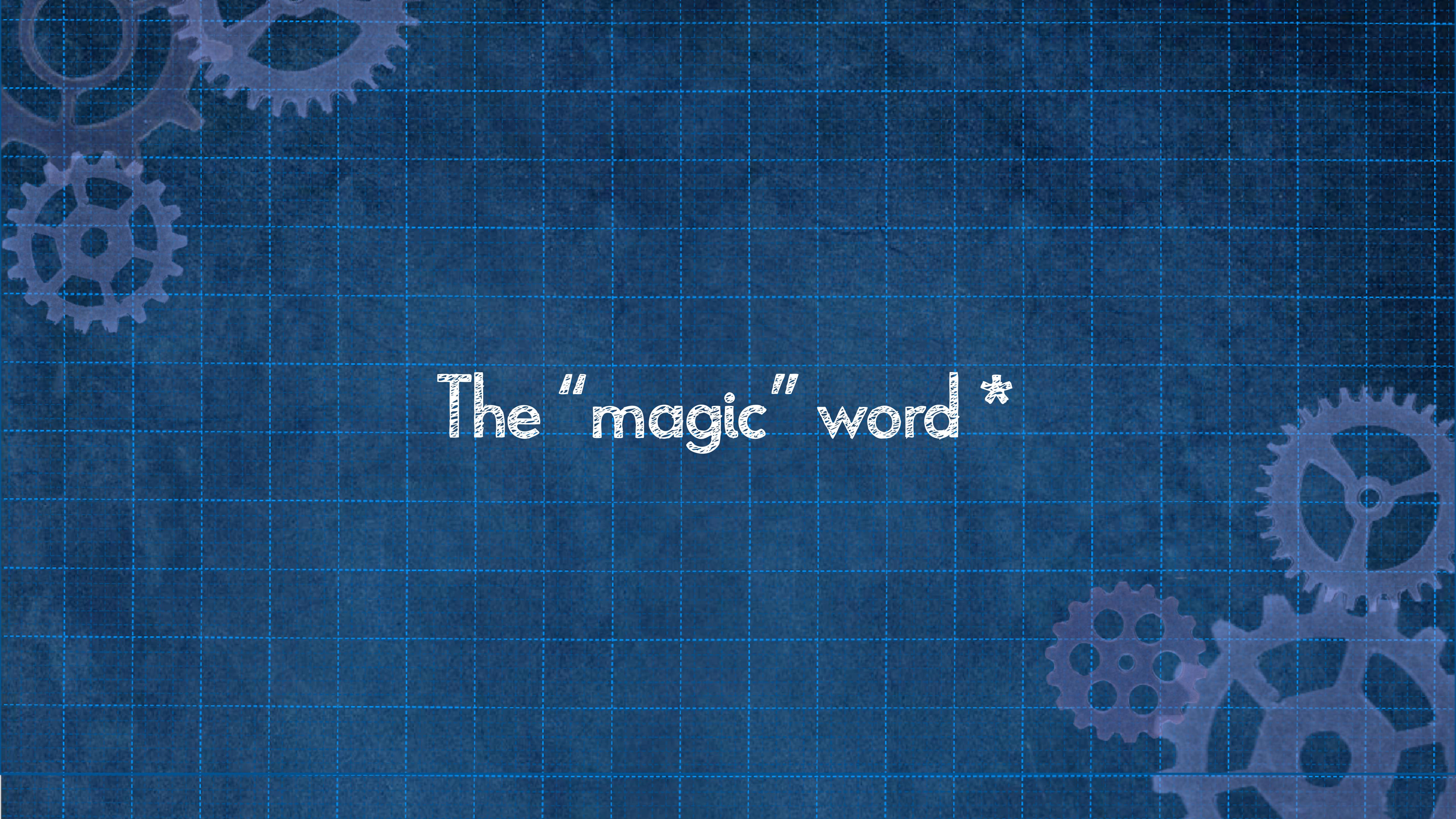
Be very clear and specific



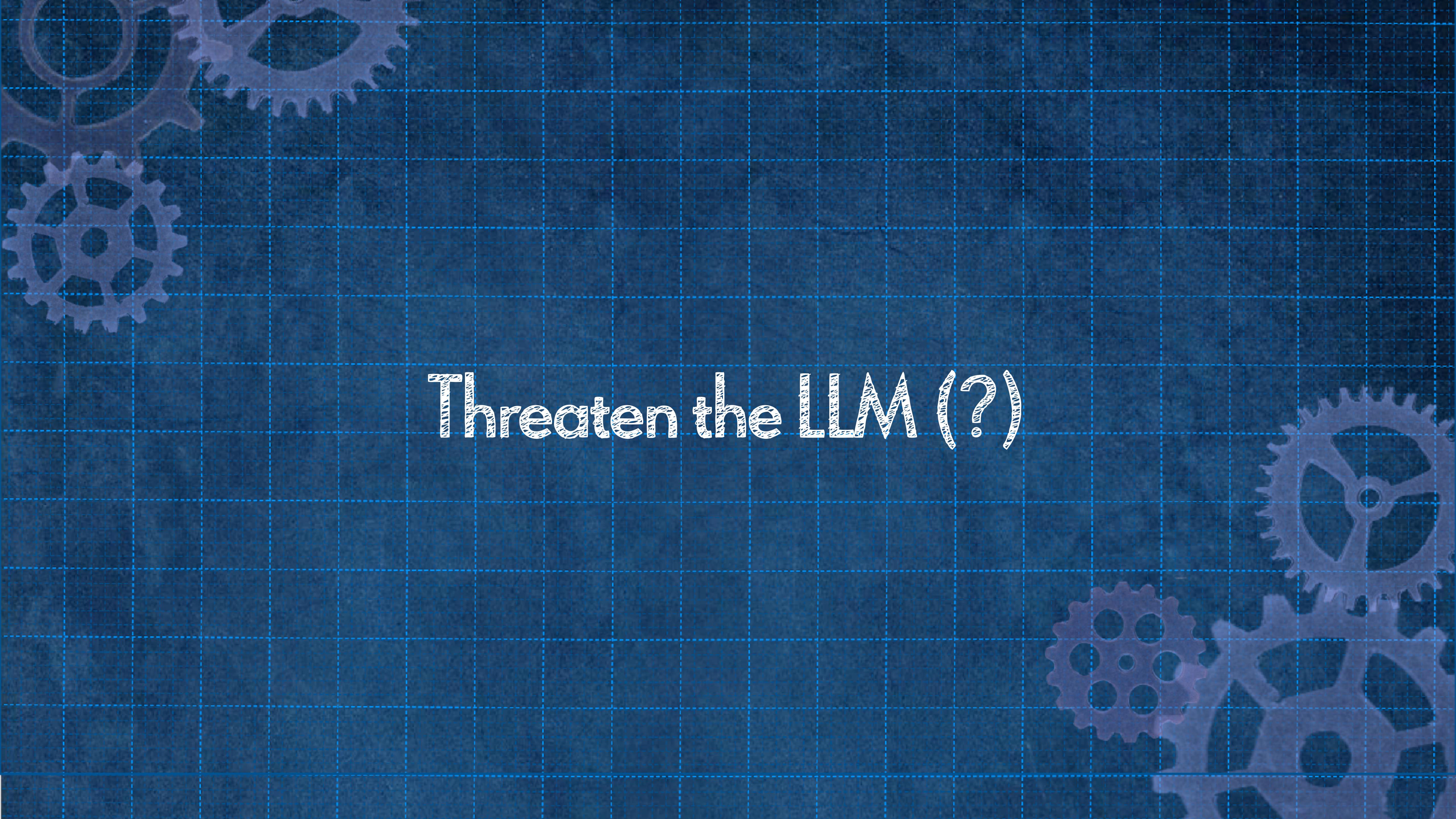
Offer structured input and ask for
structured output



Assign a role



The "magic" word *



Threaten the LLM (?)

Setting prompt options

Starting with the phrase “I have a large collection of”

The model is consulted to pick the next token

(But it's hard to think in terms of tokens; words are easier)

It comes up with 5 candidates...

books, coins, records, arts, stamps

Each has a “logit”

E.g., a relative score of how confident the LLM thinks that
should be the next token

LOGITS

books

coins

records

arts

stamps

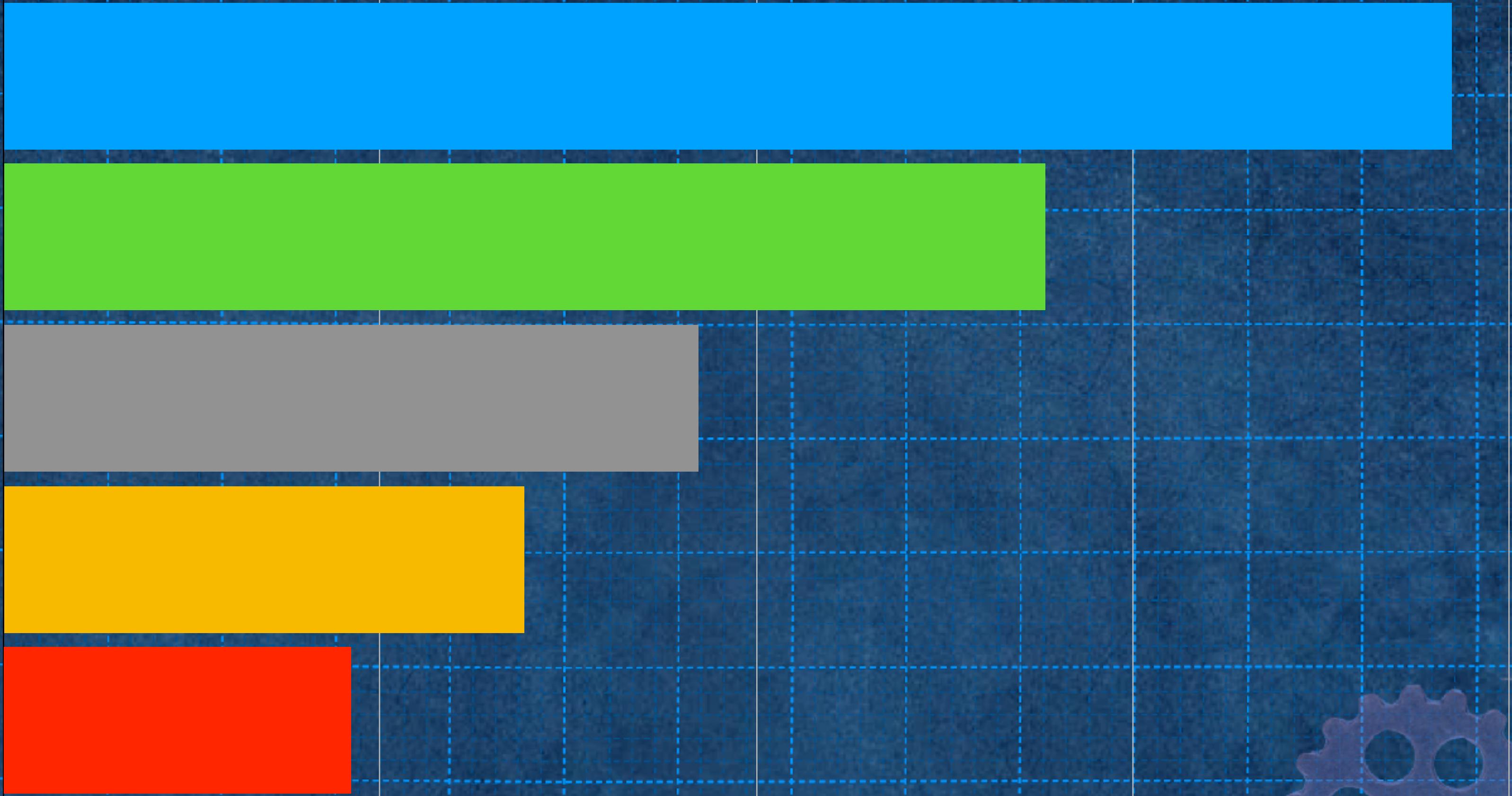
0

0.65

1.3

1.95

2.6



SOFTMAX → PROBABILITIES

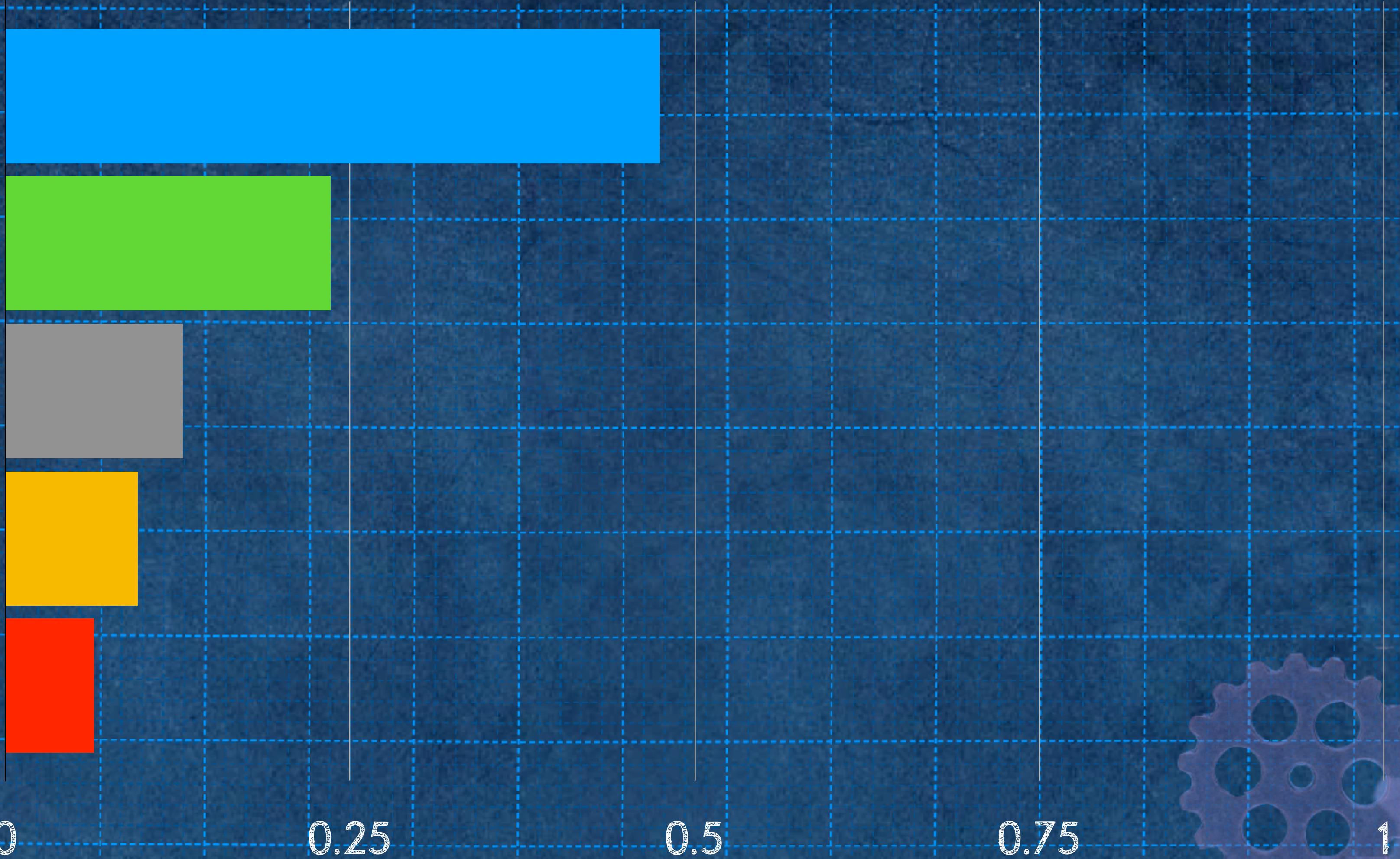
books

coins

records

arts

stamps



TEMPERATURE

books

coins

records

arts

stamps

0

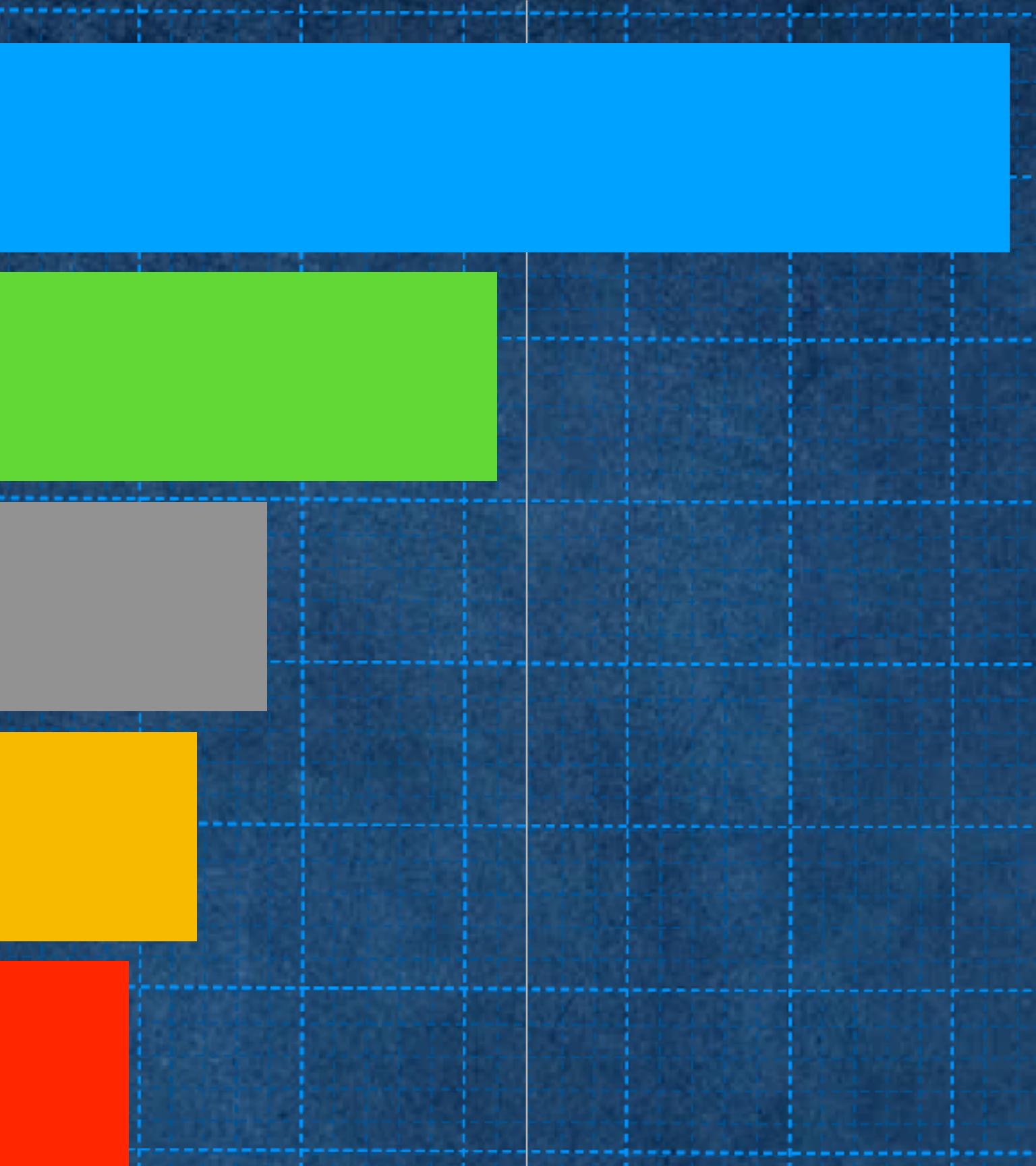
0.25

0.5

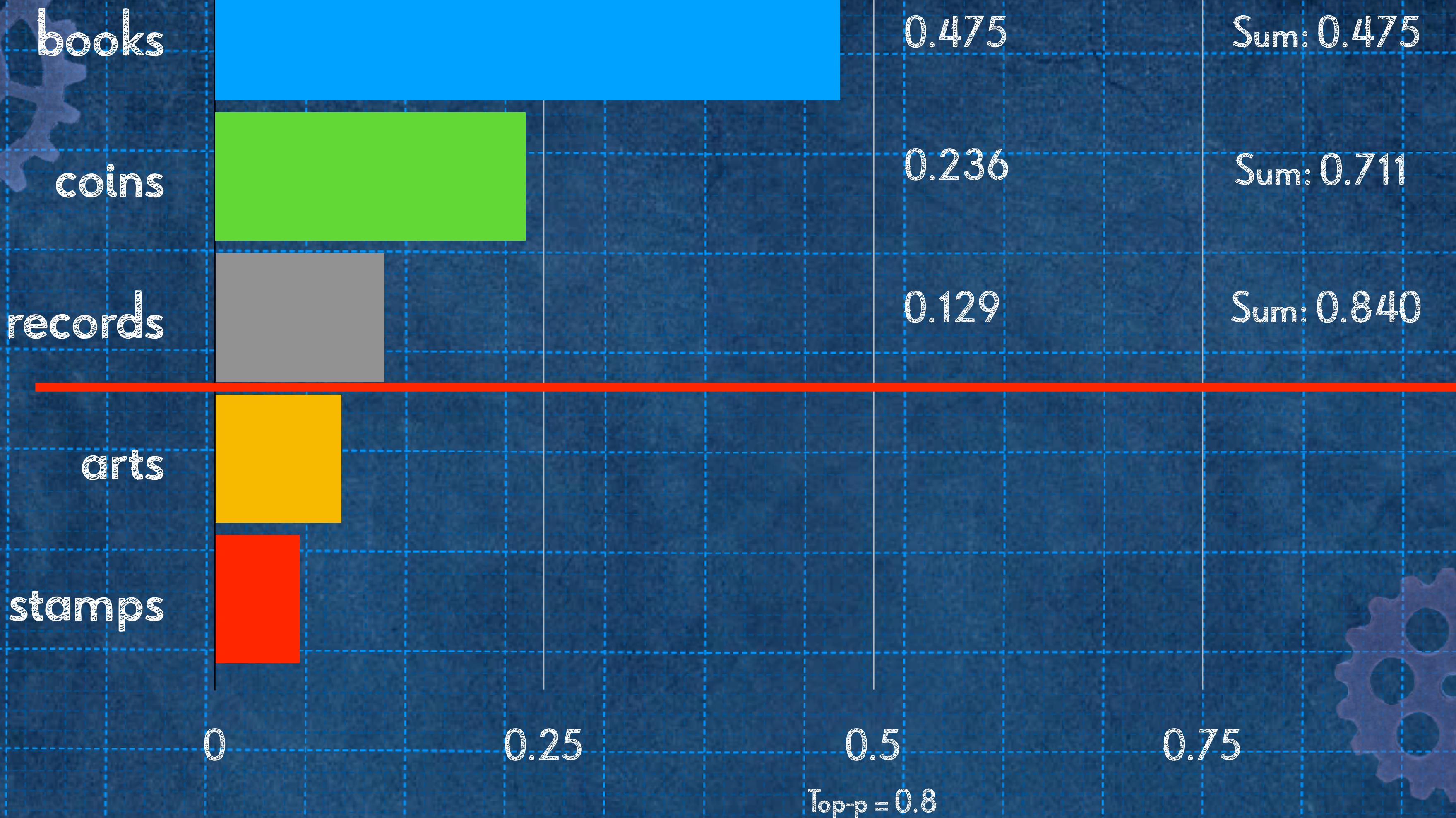
0.75

1

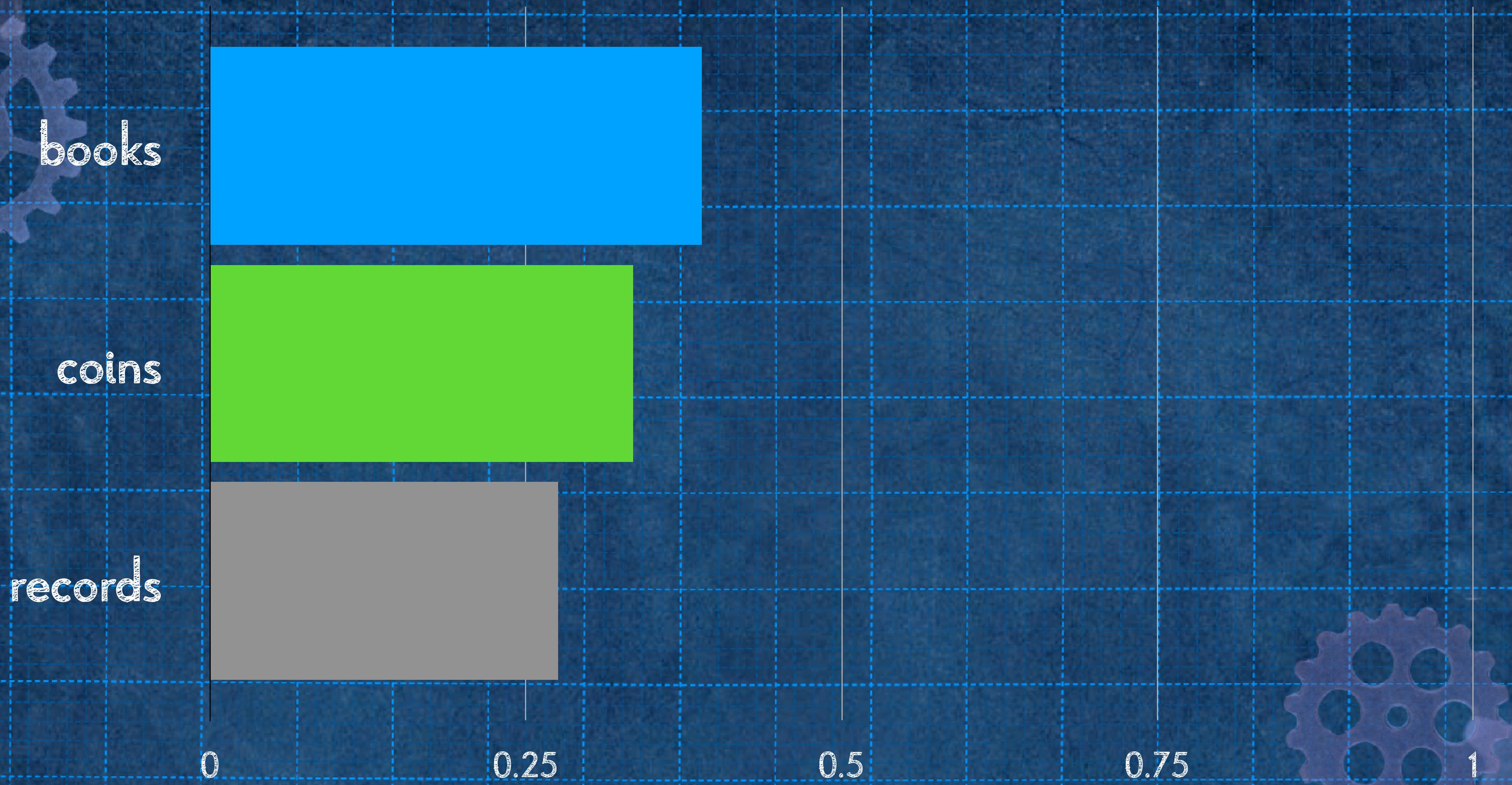
Temperature=1



TOP-P



REDISTRIBUTION AFTER PIP



TOP-K

books

coins

records

arts

stamps

0

0.25

0.5

0.75

Top-k = 2

1

REDISTRIBUTION AFTER TOP-K

books

coins

0

0.25

0.5

0.75

1

0.668

0.332



LOGITBIAS

books

coins

records

arts

stamps

0

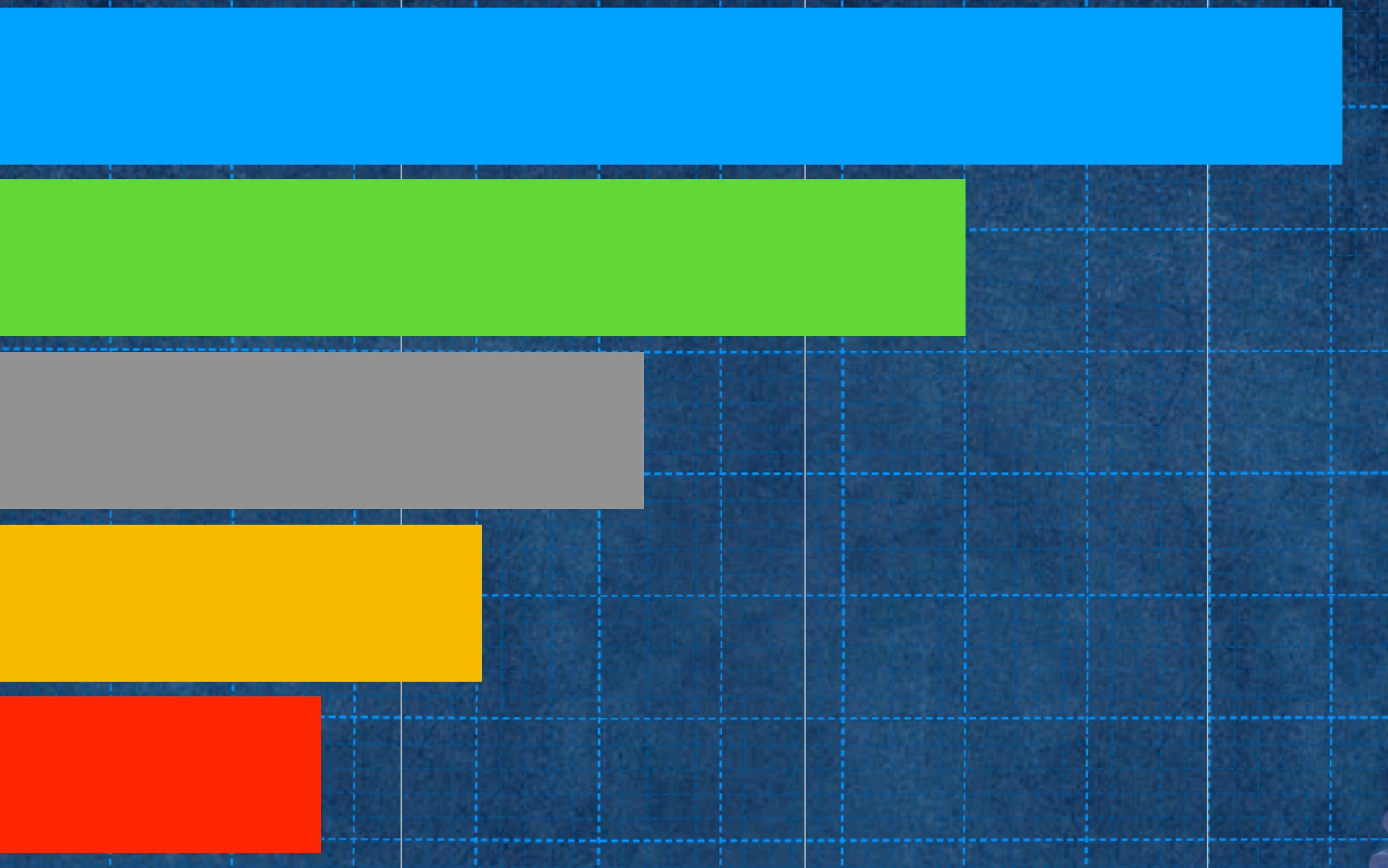
0.75

1.5

2.25

3

Base Logits



LOGITBIAS

books

coins

records

arts

stamps

0

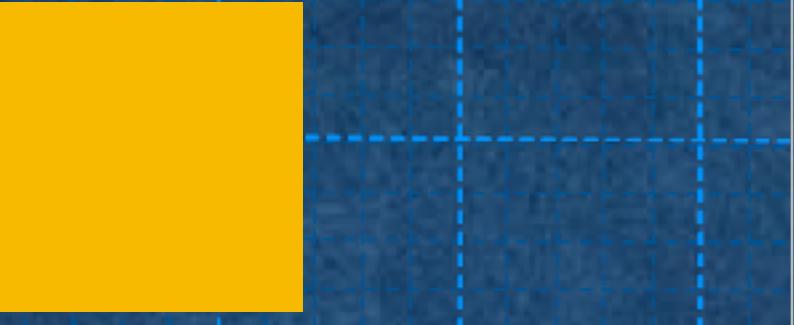
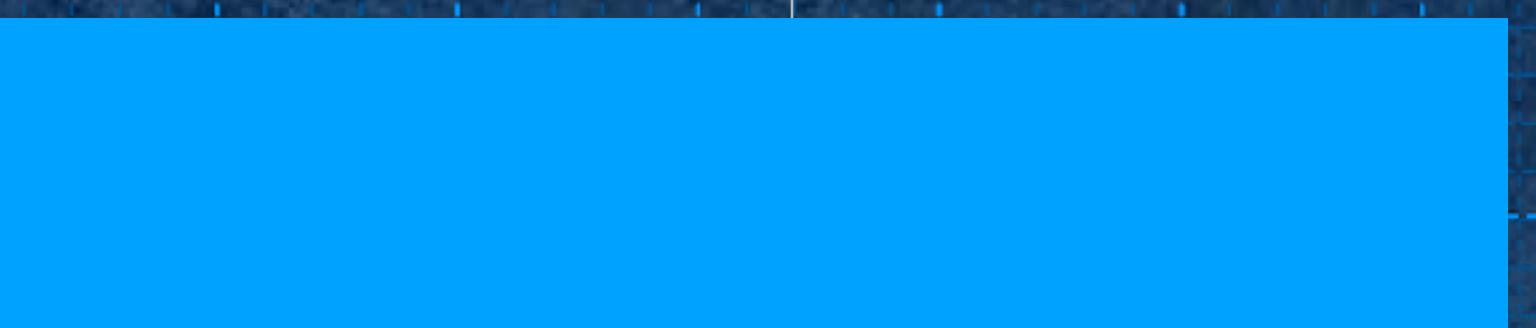
0.25

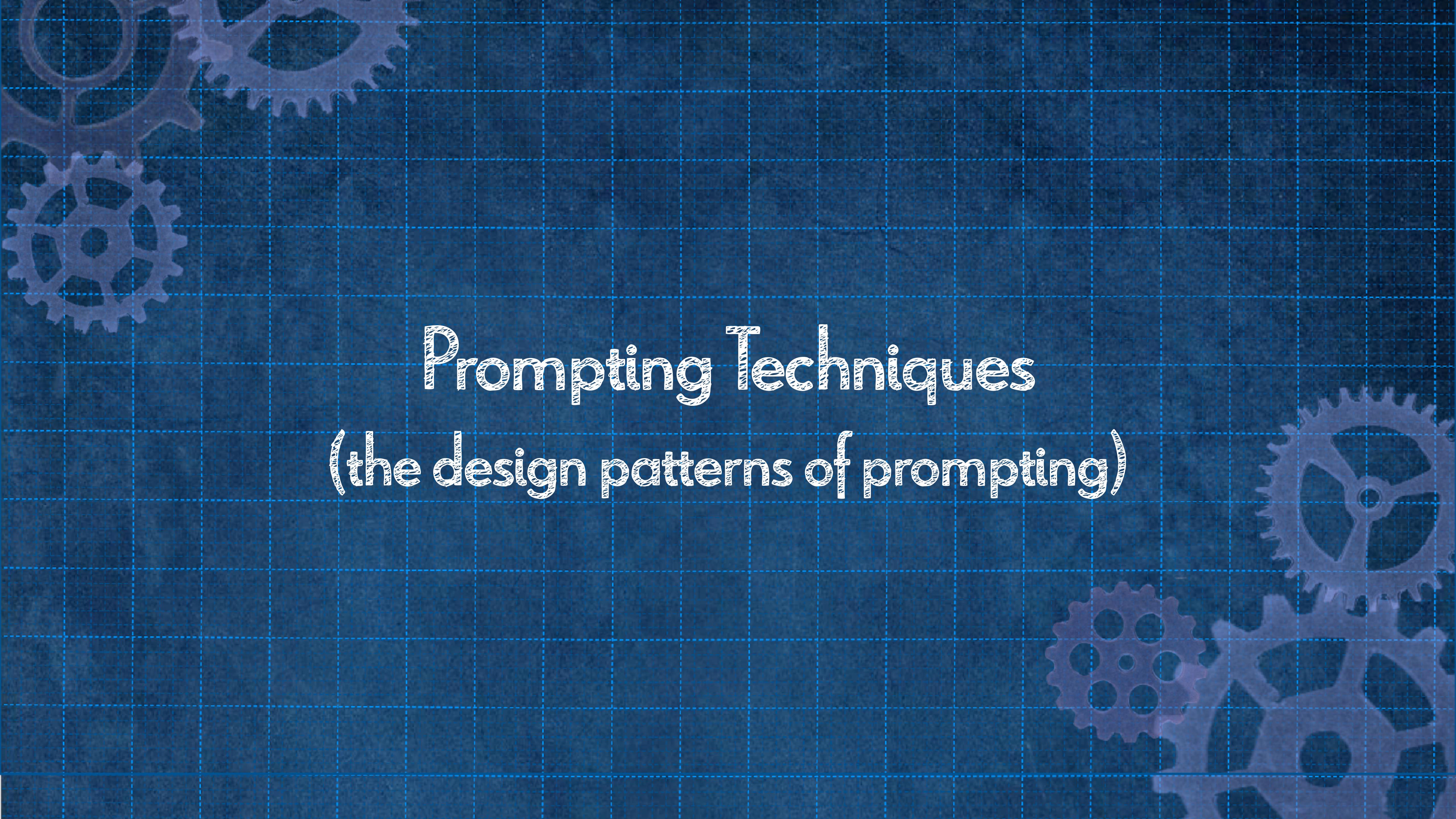
0.5

0.75

Logit-bias=0

1



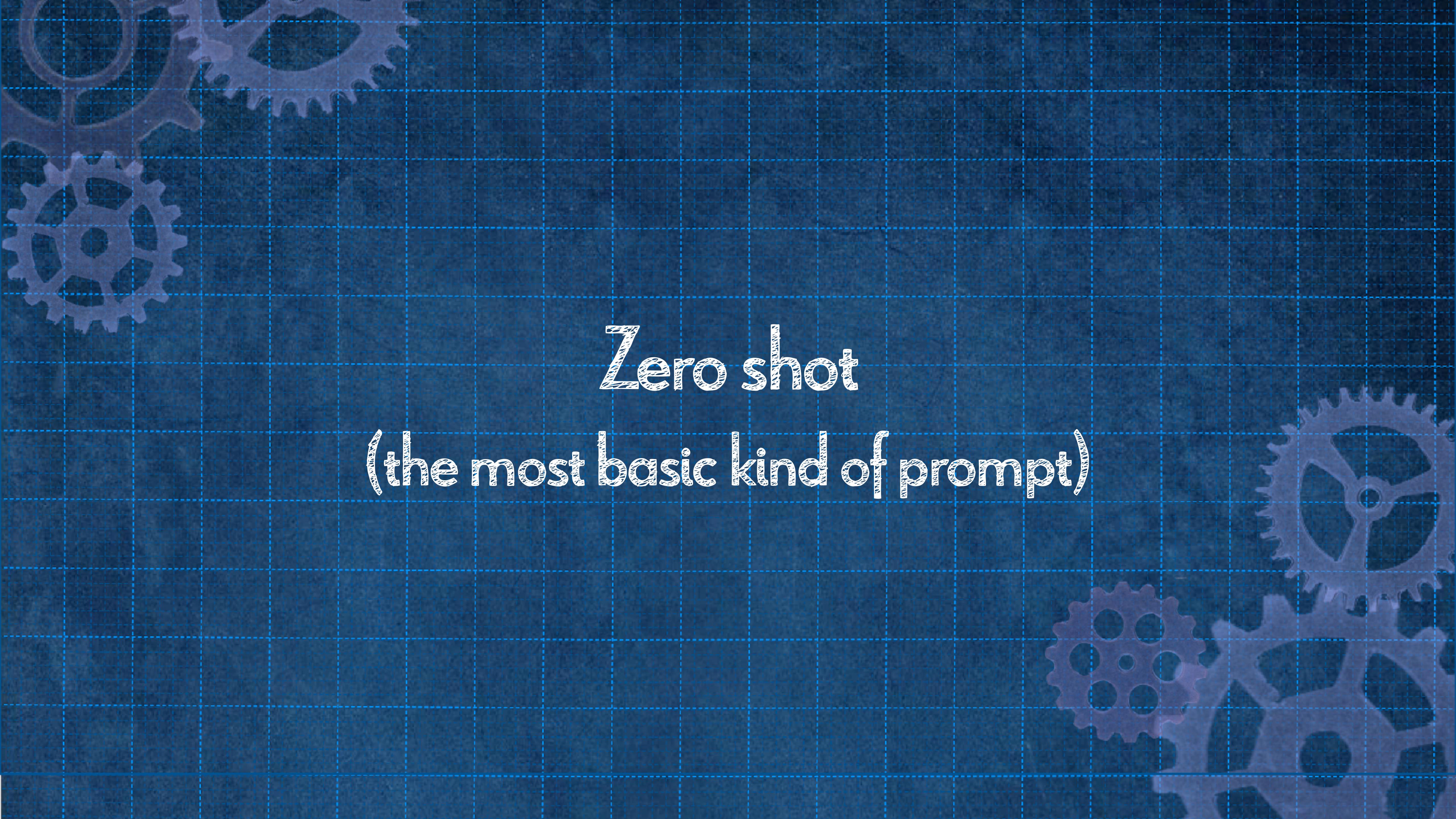


Prompting Techniques

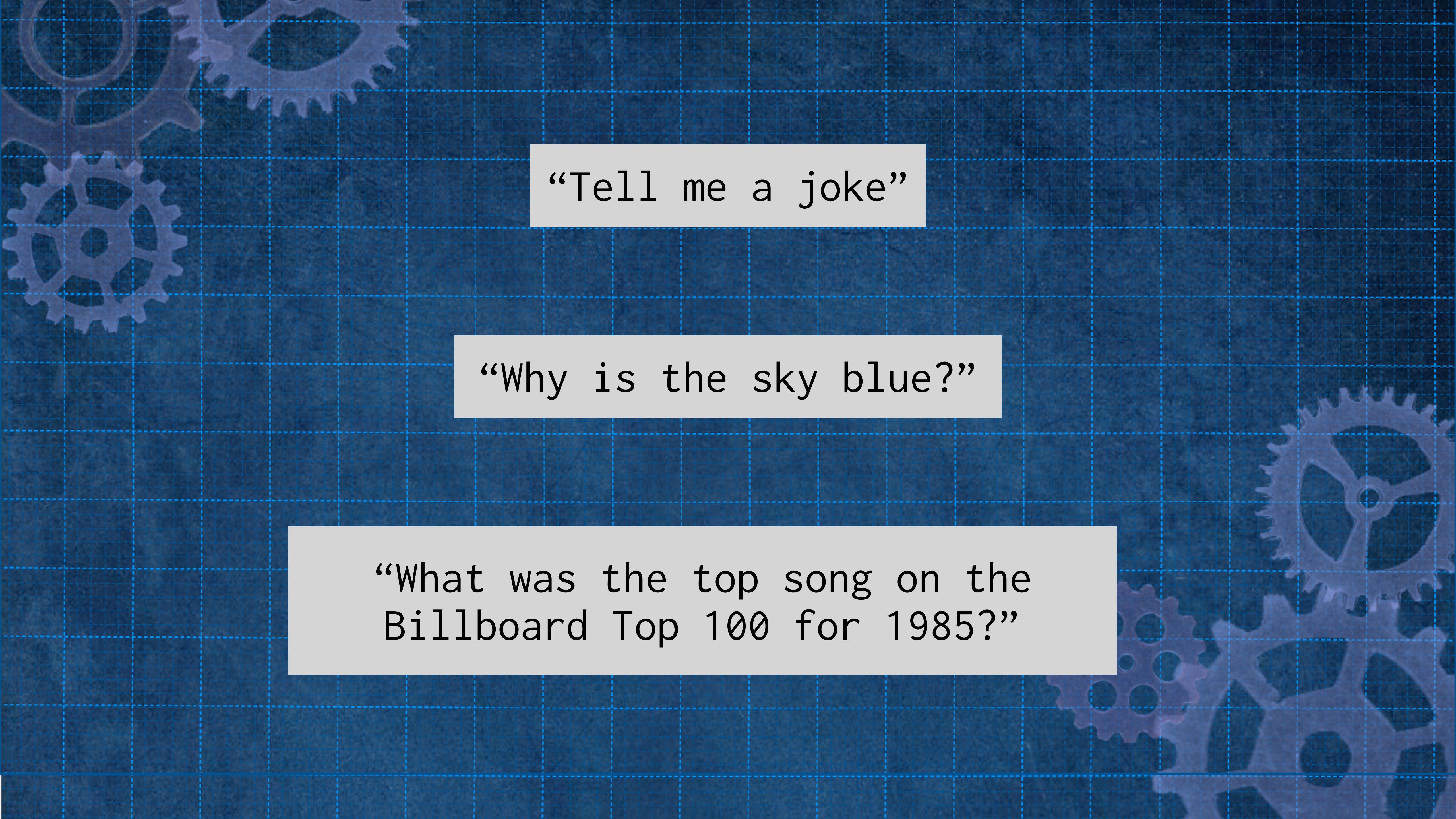
(the design patterns of prompting)

See also...

- <https://promptingguide.ai>
- <https://platform.openai.com/docs/guides/text>
- <https://docs.anthropic.com/en/docs/build-with-claude/prompt-engineering/overview>
- <https://workspace.google.com/learning/content/gemini-prompt-guide>



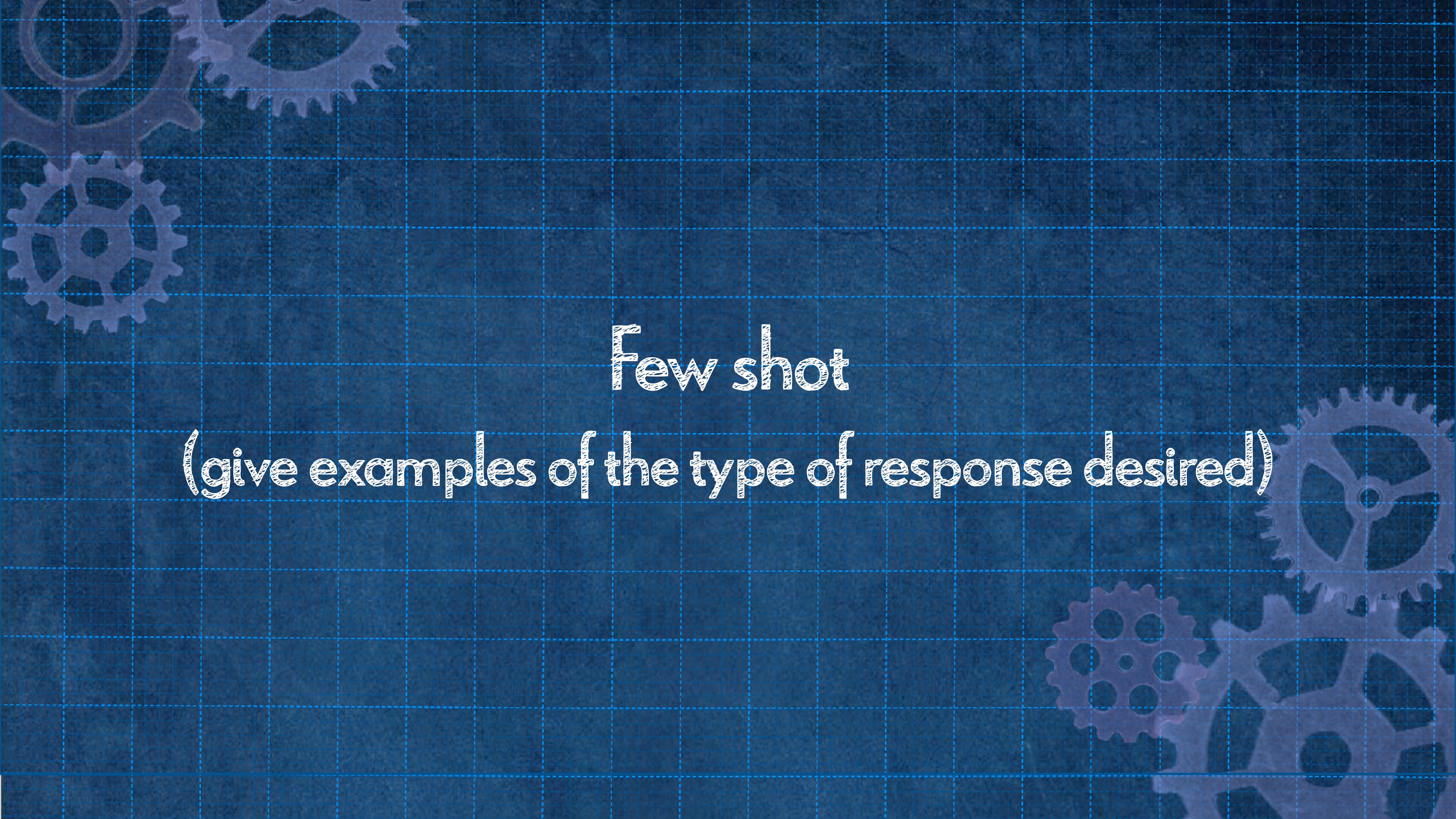
Zero shot
(the most basic kind of prompt)



“Tell me a joke”

“Why is the sky blue?”

“What was the top song on the
Billboard Top 100 for 1985?”



Few shot
(give examples of the type of response desired)

User: Hello

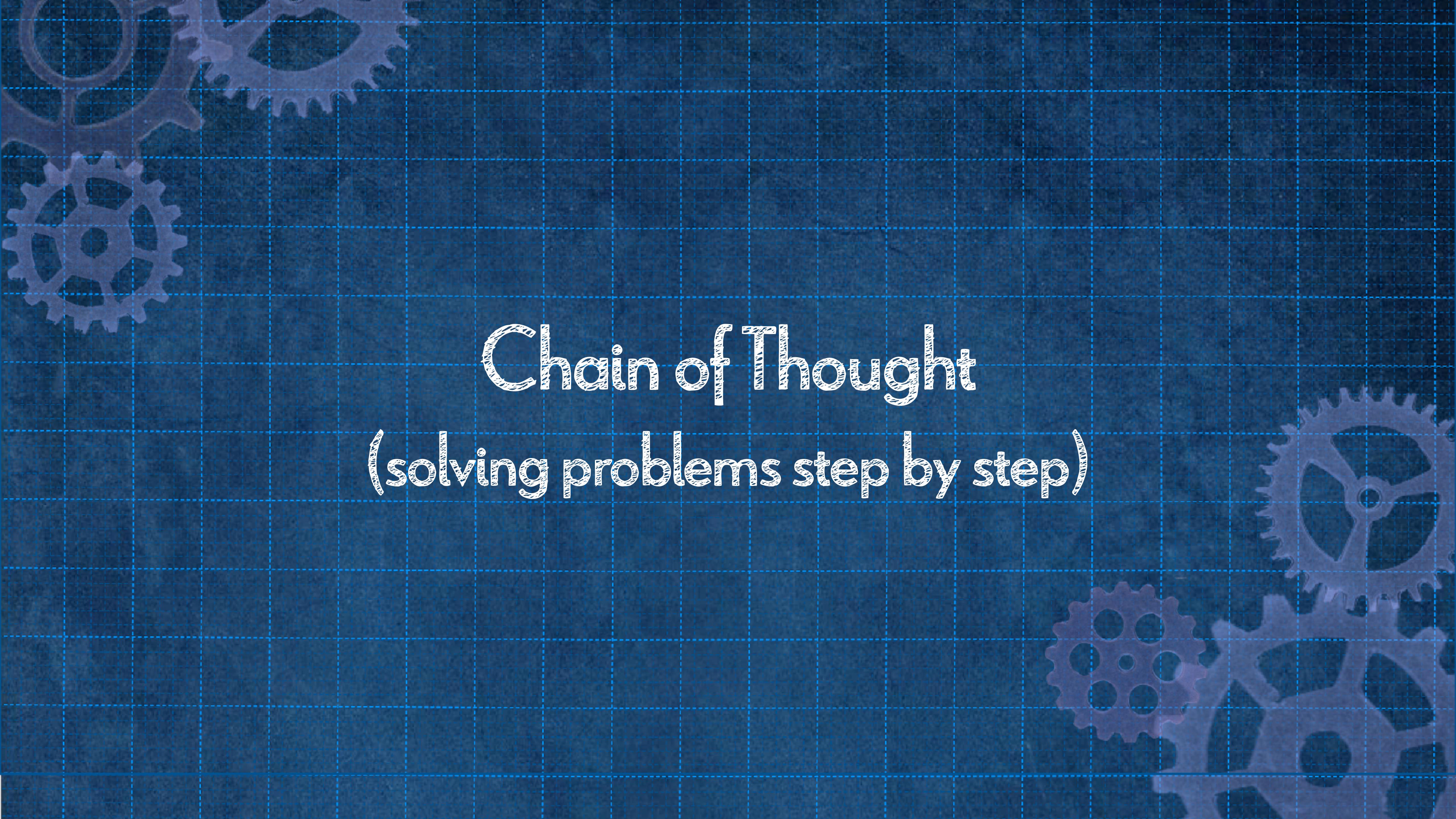
Response: Ah, greetings, my most cherished and beloved companion! How my heart swells with joy at the mere sight of you! It is as if the very sun itself shines brighter in your presence, and the birds seem to sing a sweeter melody in honor of your arrival. Pray, allow me to express the depth of my delight at this serendipitous meeting, as though the heavens themselves have conspired to bring us together in this moment. Truly, it is a gift beyond measure to be in your company, and I am overjoyed by the opportunity to share in this fleeting, yet wondrous exchange.

User: Goodbye

Response: Ah, alas, the time has come for us to part, and though my heart feels a sorrow as deep as the oceans themselves, I shall carry the memory of this moment like a treasured jewel, tucked safely within the chambers of my soul. May the winds of fortune guide your every step, and may the stars above light your path, ever bright and full of promise. Though we must now bid adieu, know that the bond we share is eternal, woven in the very fabric of time. Farewell, dear one, until we are graced with the bliss of meeting again.

User: Pass the salt

Response:



Chain of Thought
(solving problems step by step)

Chain of Thought

System prompt

Think step by step to answer the following question. Return the answer at the end of the response after a separator #####.

User prompt

Which is a faster way to get to work?

Option 1: Take a 1000 minute bus, then a half hour train, and finally a 10 minute bike ride.

Option 2: Take an 800 minute bus, then an hour train, and finally a 30 minute bike ride.



Chain of Draft

(Chain of Thought with fewer details)

*<https://arxiv.org/html/2502.1860v1>

Chain of Draft

System prompt

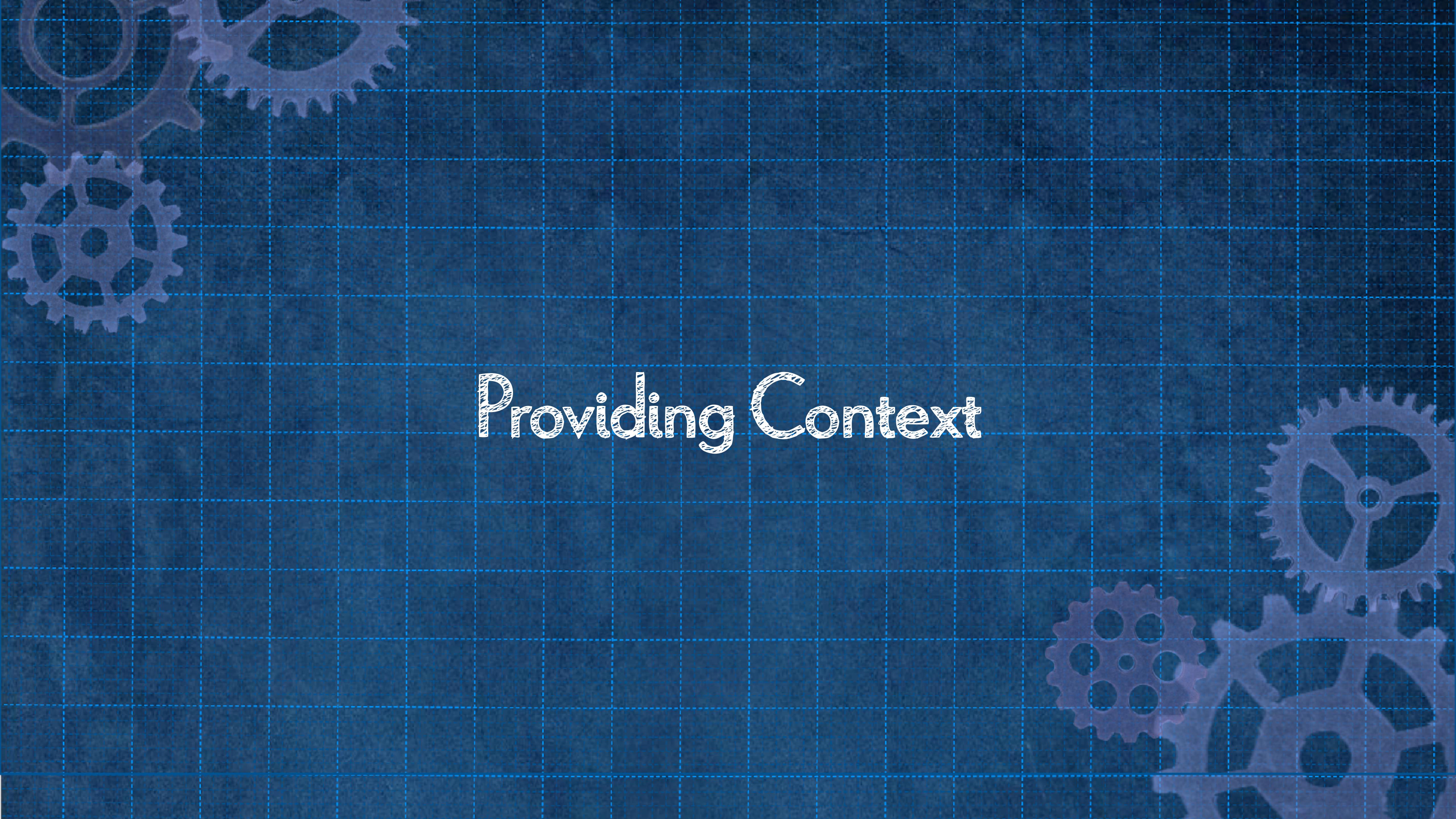
Think step by step, but only keep a minimum draft for each thinking step, with 5 words at most. Return the answer at the end of the response after a separator #####.

User prompt

Which is a faster way to get to work?

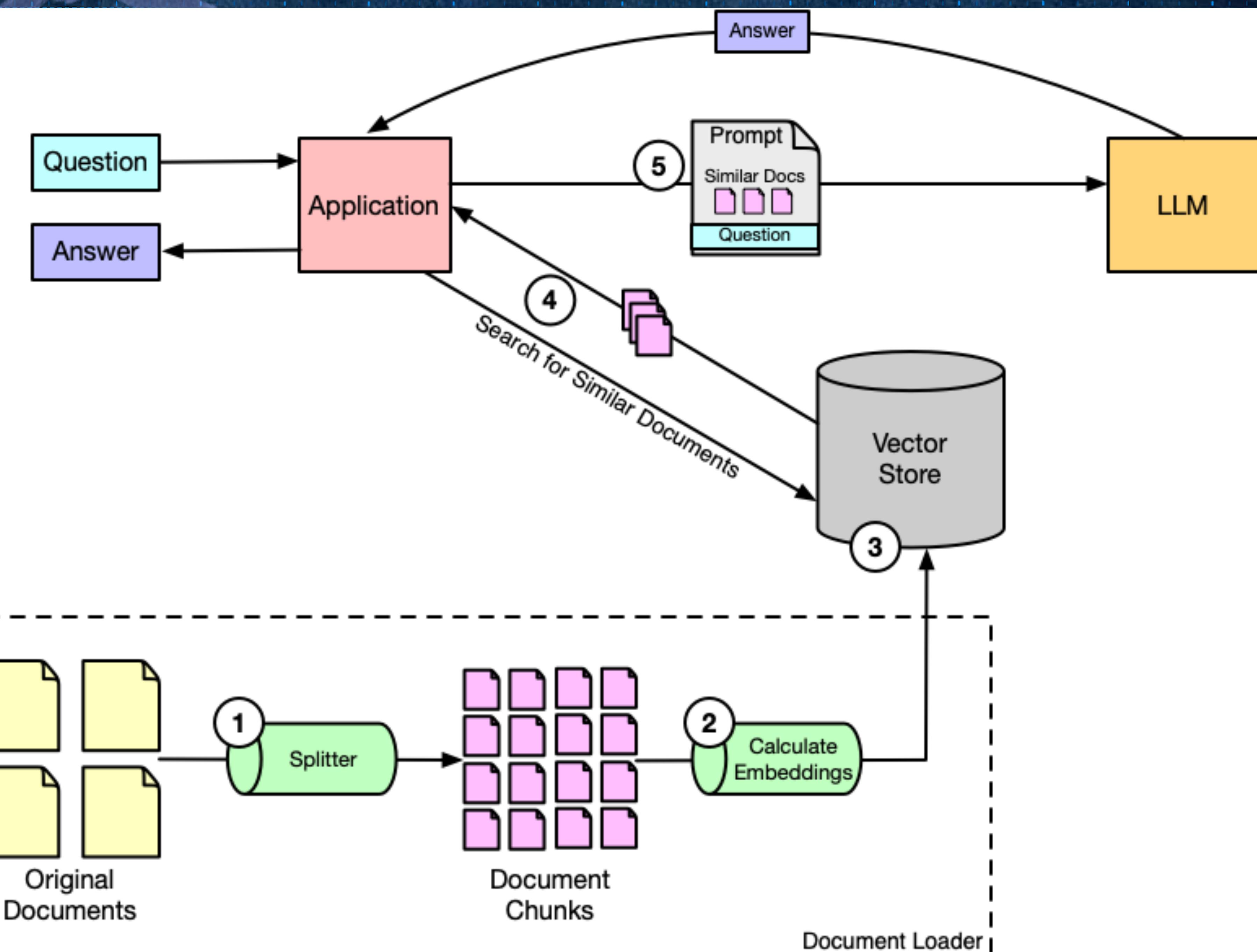
Option 1: Take a 1000 minute bus, then a half hour train, and finally a 10 minute bike ride.

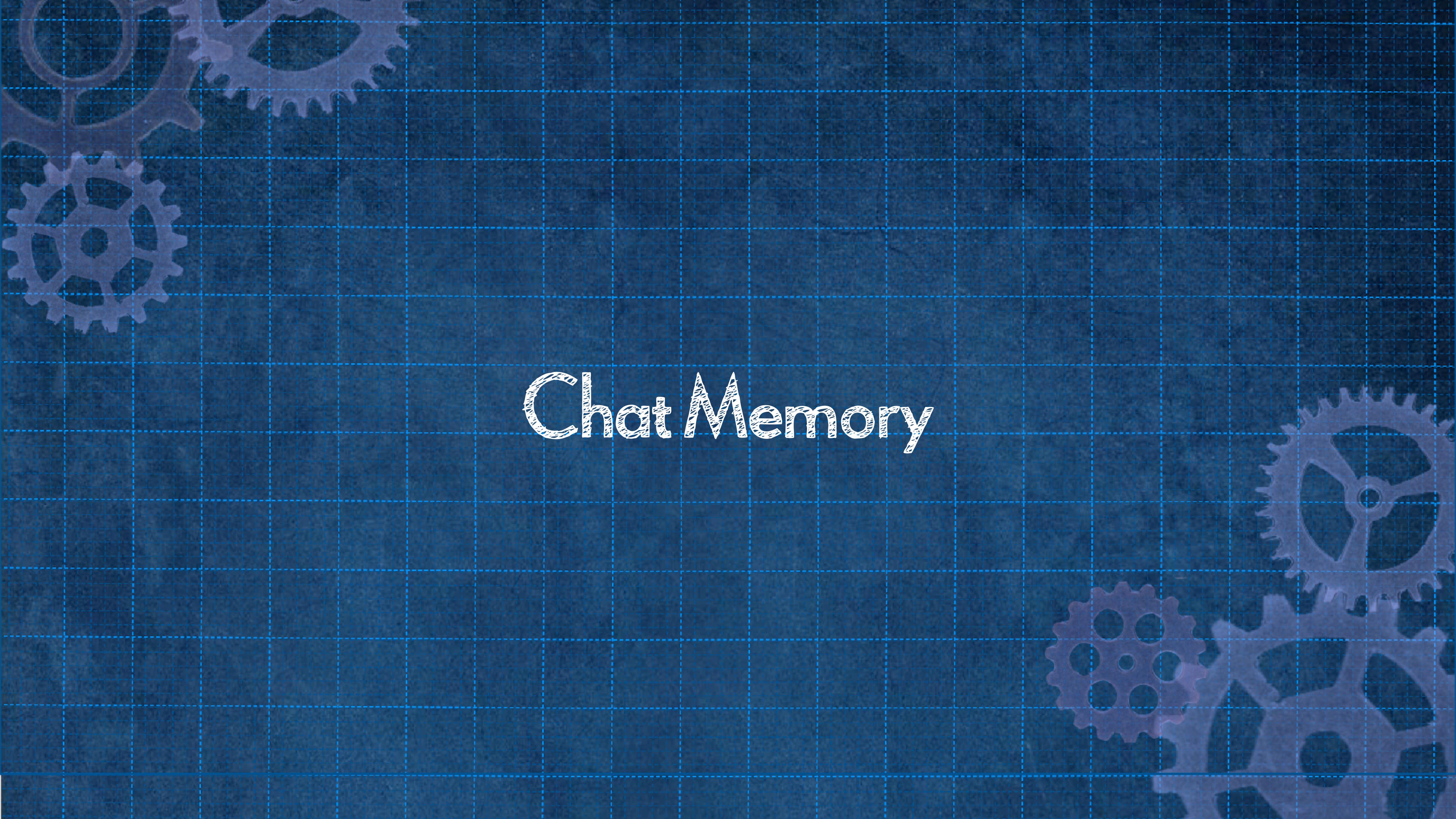
Option 2: Take an 800 minute bus, then an hour train, and finally a 30 minute bike ride.



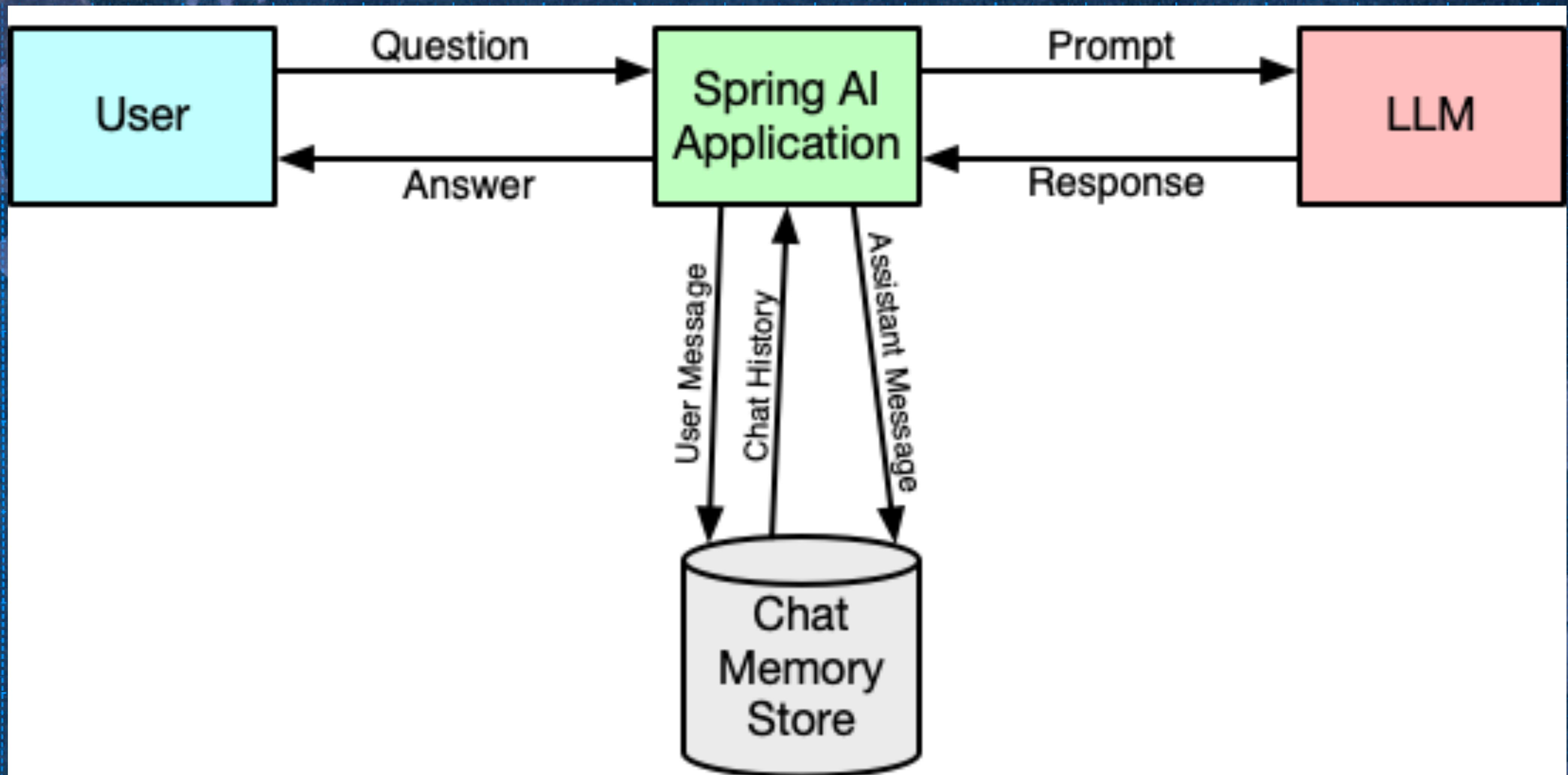
Providing Context

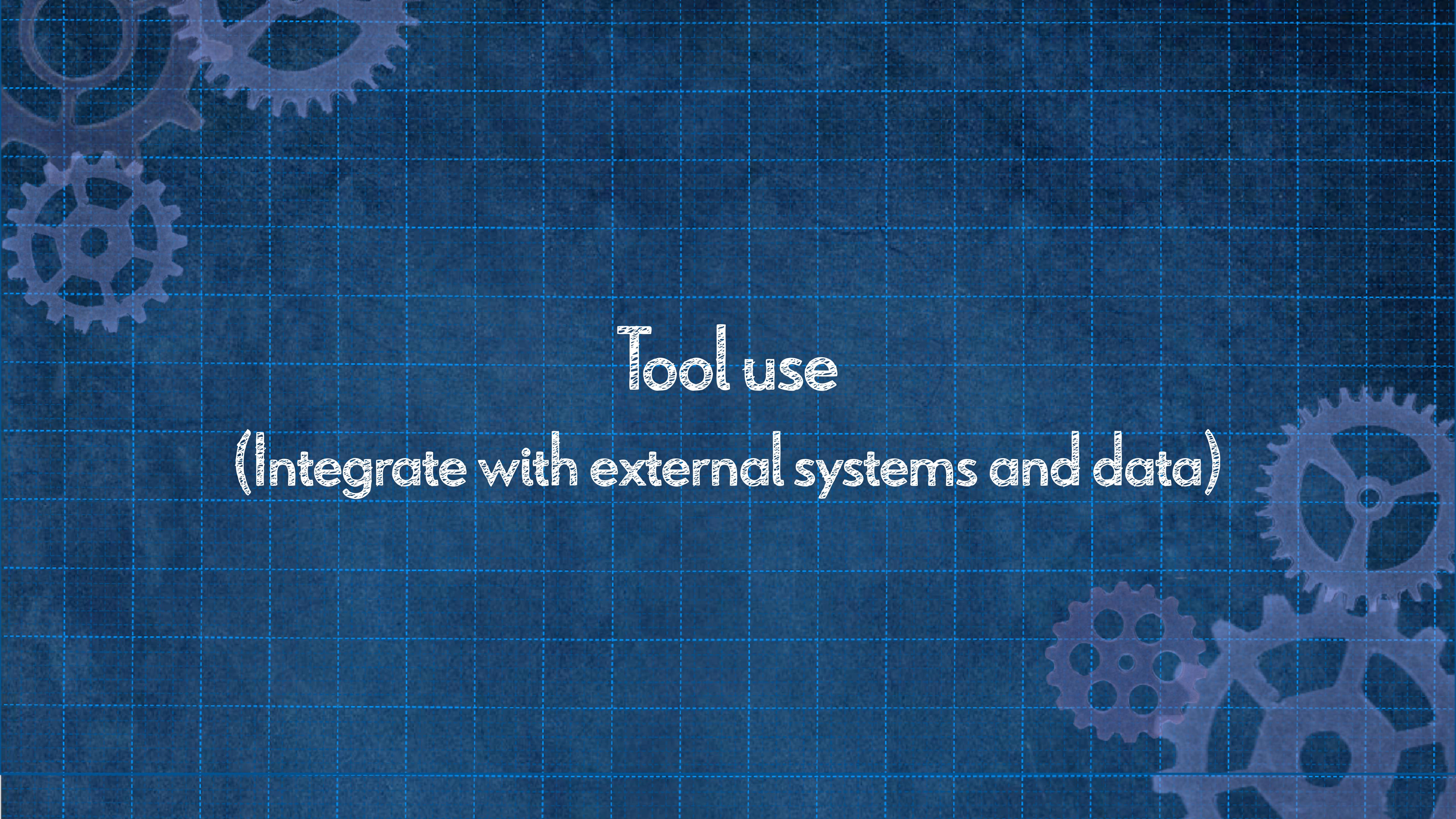
Retrieval Augmented Generation (RAG)





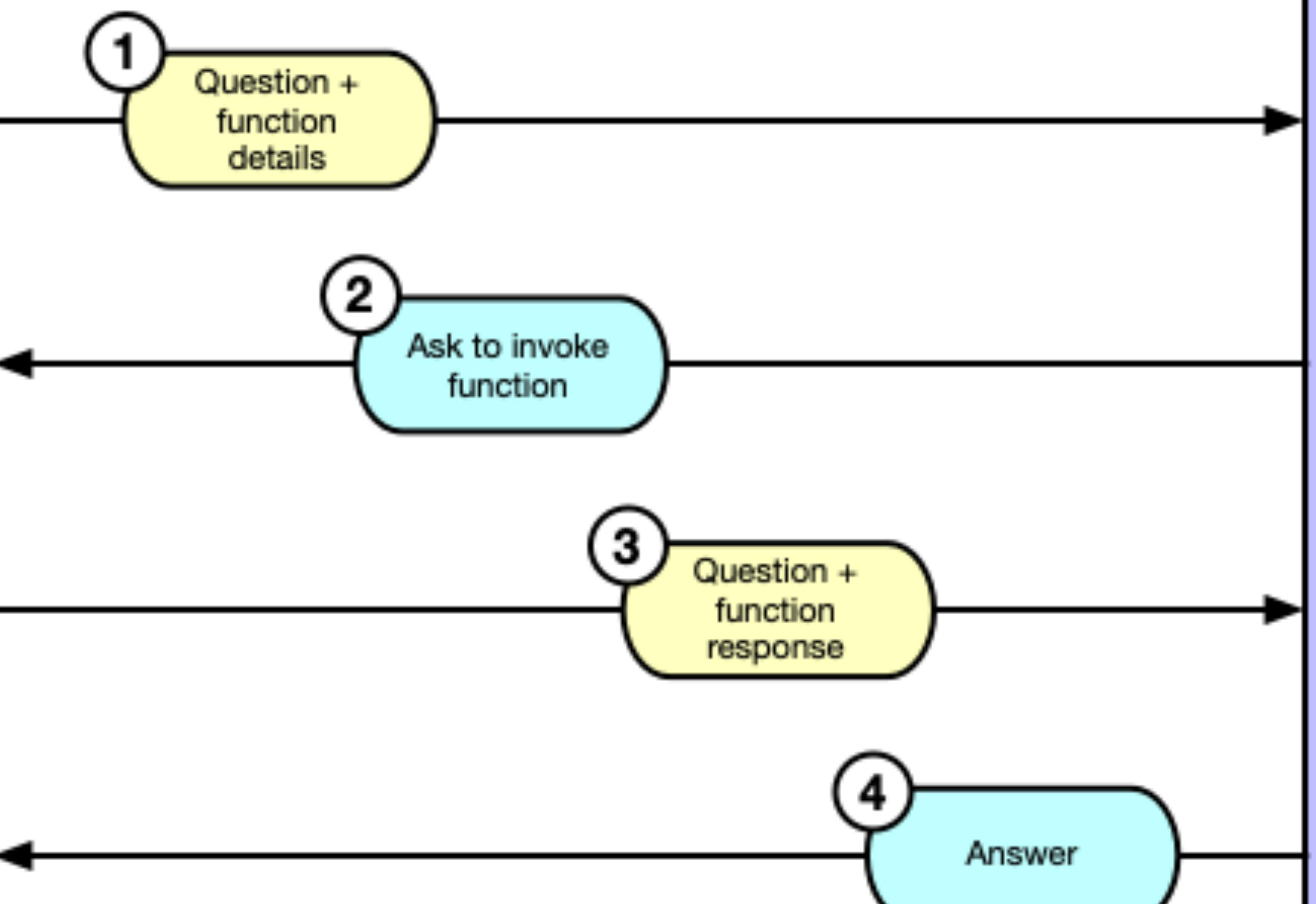
Chat Memory





Tool use
(Integrate with external systems and data)

Spring AI Application



LLM

Model Context Protocol (MCP)

<https://modelcontextprotocol.io>

What MCP offers...

Tools

Prompts

Resources

Samplings

Available MCP SDKs...

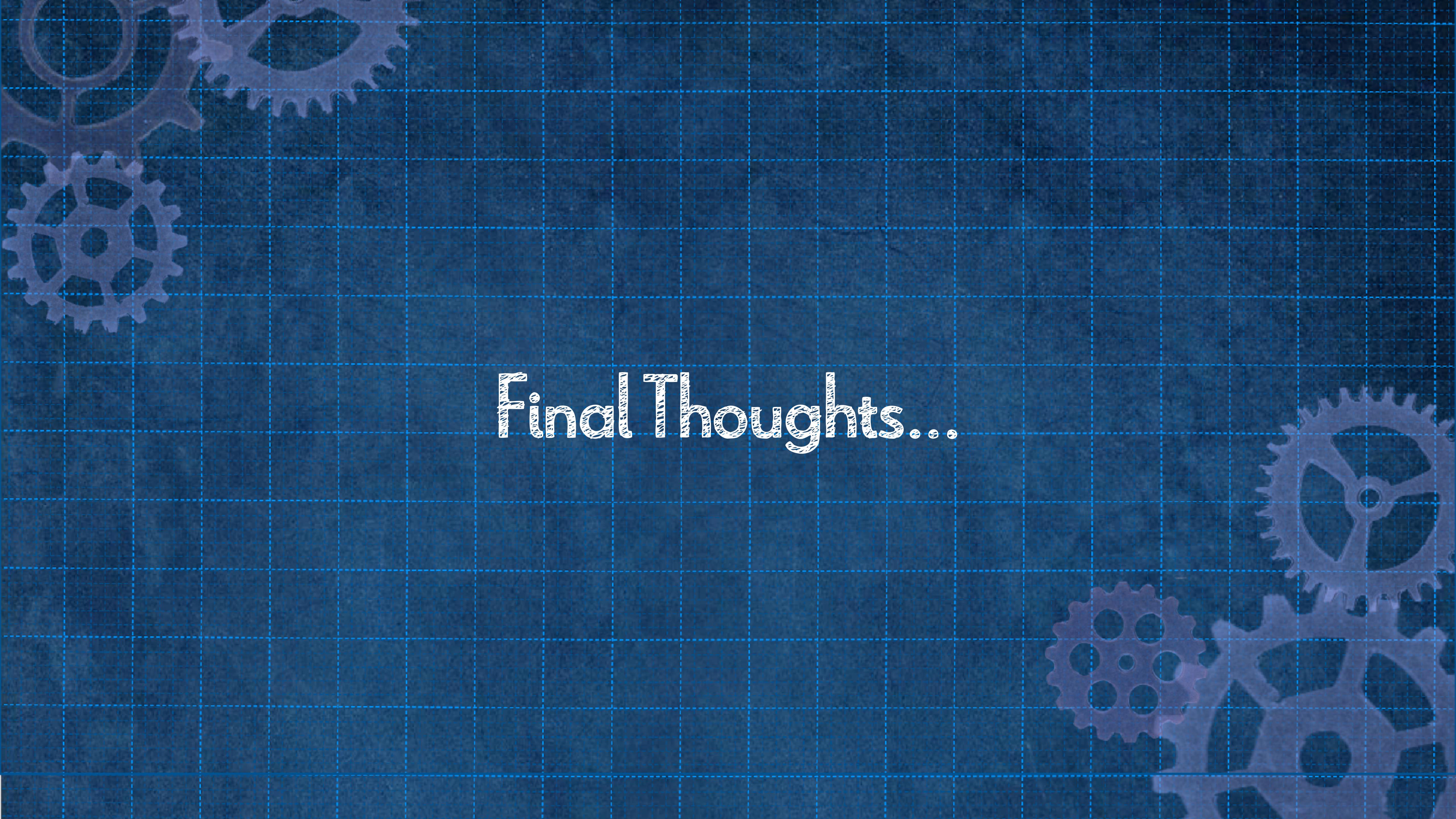
Python

TypeScript

Java

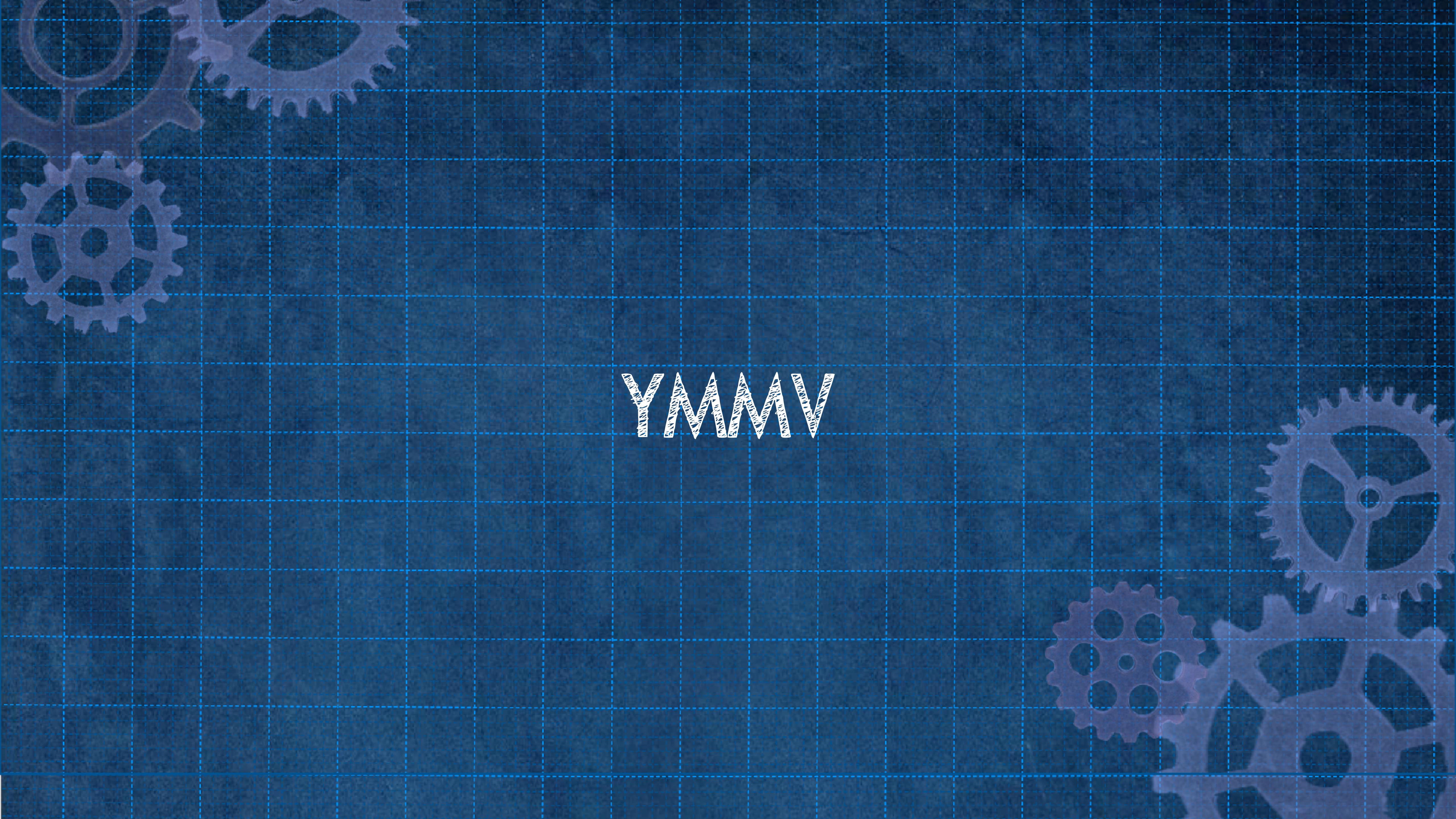
Kotlin

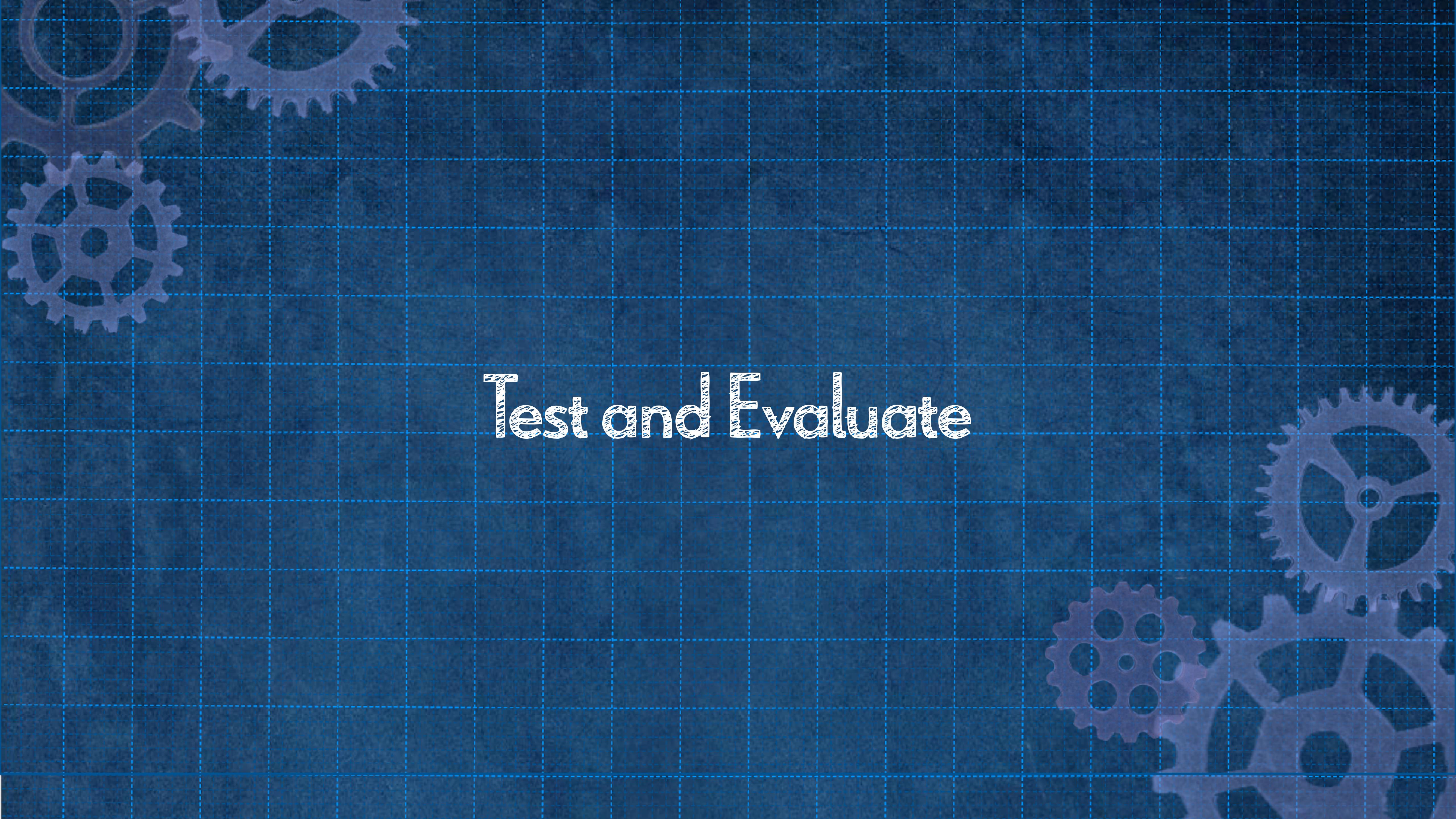
C#



Final Thoughts...

YMMV
ANONYMOUS LEADER





Test and Evaluate



Have fun!

THANK YOU

 @habuma

 @habuma

 @habuma.bsky.social

 @habuma@fosstodon.org

<https://www.habuma.com>

 craig@habuma.com

CHECK OUT MY BOOKS

