

A blue square containing the white text '03' in a large, bold, sans-serif font. To the right of the square is a yellow triangle pointing downwards.

03

# Prompt Design

# In this module, you learn to ...

- 01 Design prompts for various Gen AI use cases
- 02 Add examples and context to customize generative AI output
- 03 Experiment with parameters for model fine-tuning



# Topics

01	Understanding Prompt Design
02	Context, Examples, and Fine-Tuning Parameters
03	Prompt Design Best Practices



# Prompt design is the process of writing instructions to generative AI models

- Well-designed prompts can help to improve the accuracy, creativity, and safety of generative AI output
- Factors to consider when designing prompts
  - Desired output
  - Target audience
  - Limitations of the generative AI model

# Some tips for writing effective prompts include...

- Be clear and concise
- Use specific language
- Provide examples
- Use negative prompts to avoid unwanted output
  - E.g. "Who are the greatest NFL quarterbacks of all time? Don't include Tom Brady. "

With careful attention to prompt design, generative AI models can be used to create powerful and useful applications

# Those last couple of slides were written by Bard again



Write for me a slide that summarizes the importance of prompt design when using generative AI



View other drafts 



Sure, here is a slide that summarizes the importance of prompt design when using generative AI:

**Title:** The Importance of Prompt Design in Generative AI

**Subtitle:** How to write effective prompts that guide generative AI models to produce the desired output.

**Body:**

- Prompt design is the process of writing instructions to generative AI models.
- Well-designed prompts can help to improve the accuracy, creativity, and safety of generative AI output.
- There are a number of factors to consider when designing prompts, including the desired output, the target audience, and the limitations of the generative AI model.
- Some tips for writing effective prompts include:

Enter a prompt here



# How prompts effect output

## Prompt

## Results

Who is Linus Torvalds?

Linus Torvalds is a Finnish software engineer who is best known as the creator and lead developer of the Linux kernel, the operating system kernel at the core of most Linux distributions. He is also the creator of the Git distributed version control system.

Summarize some important dates in the life of Linus Torvalds

1969: Linus Torvalds is born in Helsinki, Finland.  
1988: Torvalds begins studying computer science at the University of Helsinki.  
1991: Torvalds creates the Linux kernel.  
1992: The first version of Linux is released to the public.  
1996: Torvalds becomes the lead developer of Linux.  
2000: Linux is adopted by IBM as the operating system for its servers.  
2007: Torvalds is named one of Time magazine's "100 Most Influential People in the World."  
2011: Torvalds receives the ACM Turing Award for his contributions to the development of Linux.  
2015: Torvalds steps down as the lead developer of Linux, but remains an active contributor to the project.

Write me a tweet to celebrate Linus Torvald's birthday

Happy birthday to Linus Torvalds, the creator of Linux! Thank you for your contributions to the open source community. #Linux #OpenSource



# Be specific when writing your prompts

## Prompt

Write for me a short 2-level outline in bullets for a 1-page paper about Linus Torvalds

## Response

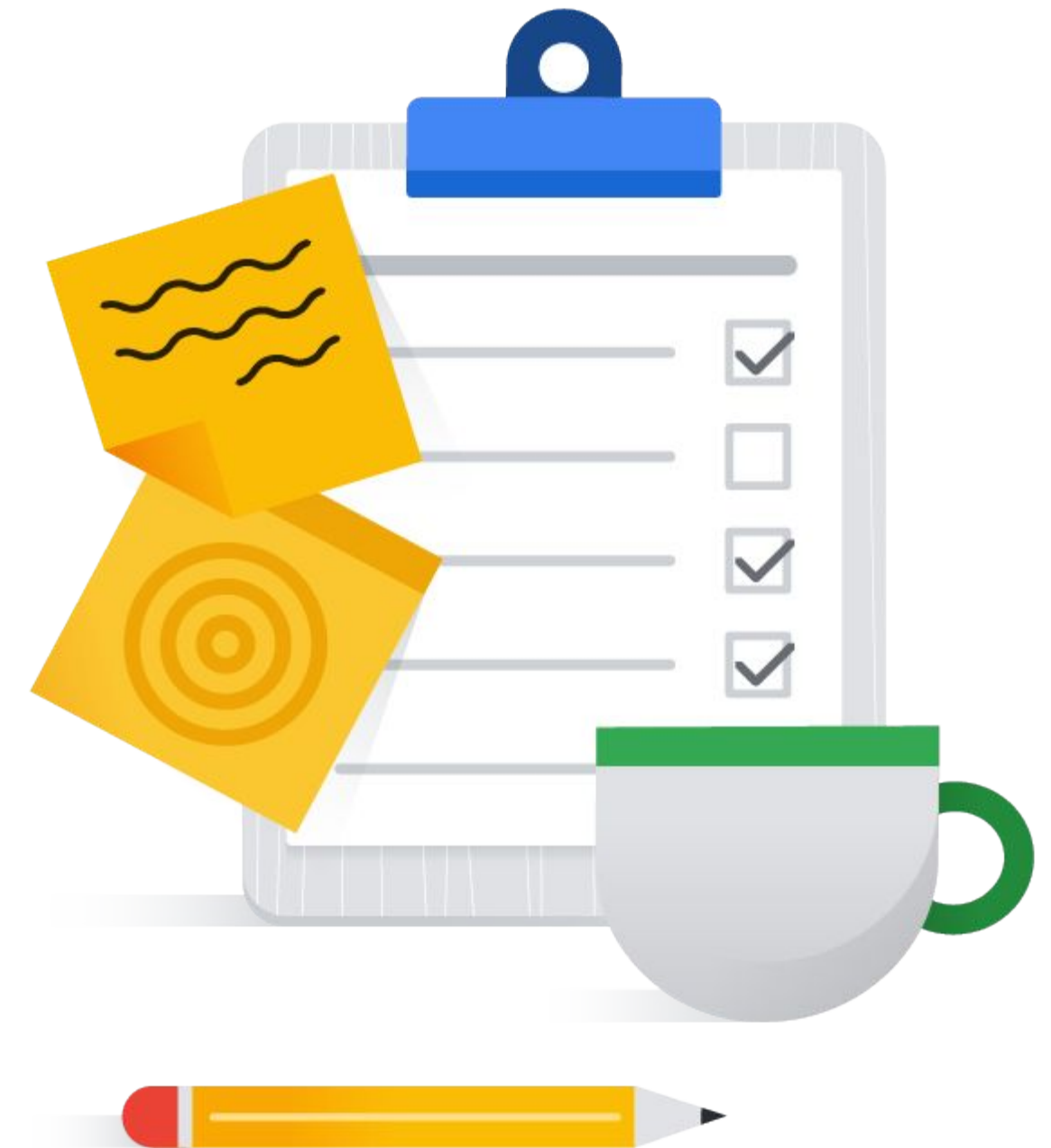
Linus Torvalds

- Biography
  - Born in Helsinki, Finland in 1969
  - Studied computer science at the University of Helsinki
  - Started working on Linux in 1991
  - Linus Torvalds is the creator of the Linux kernel, the core of the Linux operating system.
- Accomplishments
  - Linux is the most popular operating system for servers and supercomputers
  - Linux is also used on a wide variety of devices, including smartphones, tablets, and personal computers
  - Linus Torvalds has received numerous awards for his work on Linux, including the ACM Turing Award in 2015
- Influence
  - Linus Torvalds has had a profound influence on the open source software movement
  - Linux has inspired the development of many other open source operating systems, such as FreeBSD and Solaris
  - Linus Torvalds is a role model for many programmers and engineers
- ...



# Topics

01	Understanding Prompt Design
02	Context, Examples, and Fine-Tuning Parameters
03	Prompt Design Best Practices



# Quality of the output can be improved by providing context to the prompt

- Context can be a few words or a paragraph
- You can specify information about you or your organization
- Specify a genre or style for the results

# Here are the results of a prompt with no context

Context (Optional) ?

Write your context here...

Examples (Optional) ?

USER + Write an input to add a new row

AI + Write an output to add a new row

Responses

USER

Tell me about New Orleans

AI

New Orleans is a city in the U.S. state of Louisiana. It is the most populous city in Louisiana, the second-most populous city in the Southern United States, and the 41st-most populous city in the United States. New Orleans is the parish seat of Orleans Parish, the most populous parish in Louisiana. The city is located on the Mississippi River, at the mouth of the Gulf

# Let's see how adding context changes the output

Context (Optional) ?

You work for an online travel agent selling exclusive tourist packages to popular destinations

Examples (Optional) ?

USER + Write an input to add a new row

AI + Write an output to add a new row

Responses

by the United States as part of the Louisiana Purchase.

New Orleans is a major tourist destination, known for its French Quarter, jazz music, and Mardi Gras celebrations. The city is also home to a number of museums, including the National World War II Museum and the New Orleans Museum of Art.

If you're looking for a unique travel experience, New Orleans is the perfect destination. The city has something to offer everyone, from history buffs to music lovers to

# Let's try a different context

Context (Optional) ?

You work for the National Weather service writing blog posts about climate events in various locations around the world.

Examples (Optional) ?

USER + Write an input to add a new row

AI + Write an output to add a new row



Responses

located on the Mississippi River, at the mouth of the Gulf Intracoastal Waterway, 125 miles (201 km) upriver from the Gulf of Mexico. The city is known for its distinctive architecture, vibrant culture, and rich history.

New Orleans is also known for its vulnerability to hurricanes and flooding. The city has been hit by several major hurricanes in recent years, including Hurricane Katrina in 2005 and Hurricane Ida in 2021. These hurricanes caused widespread damage and displacement, and





# Let's try requesting a style of output

 **Context** (Optional) 

You work for an online travel agent that sells vacation packages to popular destinations. You write ads in the style of Dr. Seuss.

Write me an add for a vacation rental in Hawaii on the beach



AI

Oh, the places you'll go  
When you stay in Hawaii,  
On the beach, by the sea,  
With a view of the palm  
trees.

You can swim in the ocean,  
And surf on the waves, And  
build sandcastles on the  
beach, And watch the  
sunset over the water.

You can go for hikes in the  
mountains, And explore the  
rainforest, And visit the  
volcanoes, And see the  
native wildlife.

# Examples demonstrate what the output should look like

- Use a history of what you have written in the past to get the model to output content in your style
  - Zero-shot prompting means there are no examples
  - One-shot prompting means there is one example
  - Few-shot prompting means there are multiple examples



# Let's add some context and some examples and see how it affects output

Examples (Optional) ?

USER

Write ad ad for a beech vacation

AI

Great deals on 1-week rentals in Tahiti. Going fast, visit us at [www.funtimes.co](http://www.funtimes.co)

Context (Optional) ?

You work for Fun Times an online travel agent selling vacation packages to popular destinations. Your site is: [www.funtimes.com](http://www.funtimes.com)

USER

Write a social media post about

AI

The food, the shows, the fashion. Paris the world's favorite that and more.  
#funtimes

USER

Write an ad for a 1 week vacation on in Waikiki Beach.

AI

Looking for a tropical vacation? Look no further than Waikiki Beach. With its white sand beaches, crystal clear water, and endless activities, Waikiki Beach is the perfect place to relax and have fun. Book your vacation today at [www.funtimes.com](http://www.funtimes.com).

# Use the other model parameters to further customize the results

- Temperature
- Token limit
- Top-K
- Top-P

The image shows a user interface for configuring model parameters. It includes five sliders with their respective ranges and current values, a dropdown menu for the safety filter threshold, and buttons for submitting and resetting parameters. A link to report inappropriate responses is also present.

Parameter	Range	Current Value
Temperature	0 to 1	0.2
Token limit	1 to 1024	256
Top-K	1 to 40	40
Top-P	0 to 1	0.8
Safety filter threshold	Block few	0.8

[SUBMIT](#) [RESET PARAMETERS](#)

[REPORT INAPPROPRIATE RESPONSES](#)

Content processed through Vertex AI is

# Temperature controls the degree of randomness in token selection

- Range from 0 to 1
- Lower temperatures are good for prompts that require correctness
  - More likely to select the most expected next token
- Higher temperatures can lead to more diverse or unexpected results
  - More creative
- A temperature of 0 is deterministic
  - The highest probability token is always selected
  - Will always return the same result for a given prompt
- Default is a temperature of .2
  - Start there and experiment with your results

The screenshot shows a user interface for configuring AI parameters. The 'Temperature' slider is highlighted with a red box, showing a value of 0.2. Other parameters include 'Token limit' (256), 'Top-K' (40), and 'Top-P' (0.8). A 'Safety filter threshold' dropdown is set to 'Block few'. At the bottom, there are buttons for 'SUBMIT', 'RESET PARAMETERS', and a link to 'REPORT INAPPROPRIATE RESPONSES'. A footer note states 'Content processed through Vertex AI is'.

Parameter	Range	Current Value
Temperature	0 to 1	0.2
Token limit	1 to 1024	256
Top-K	1 to 40	40
Top-P	0 to 1	0.8

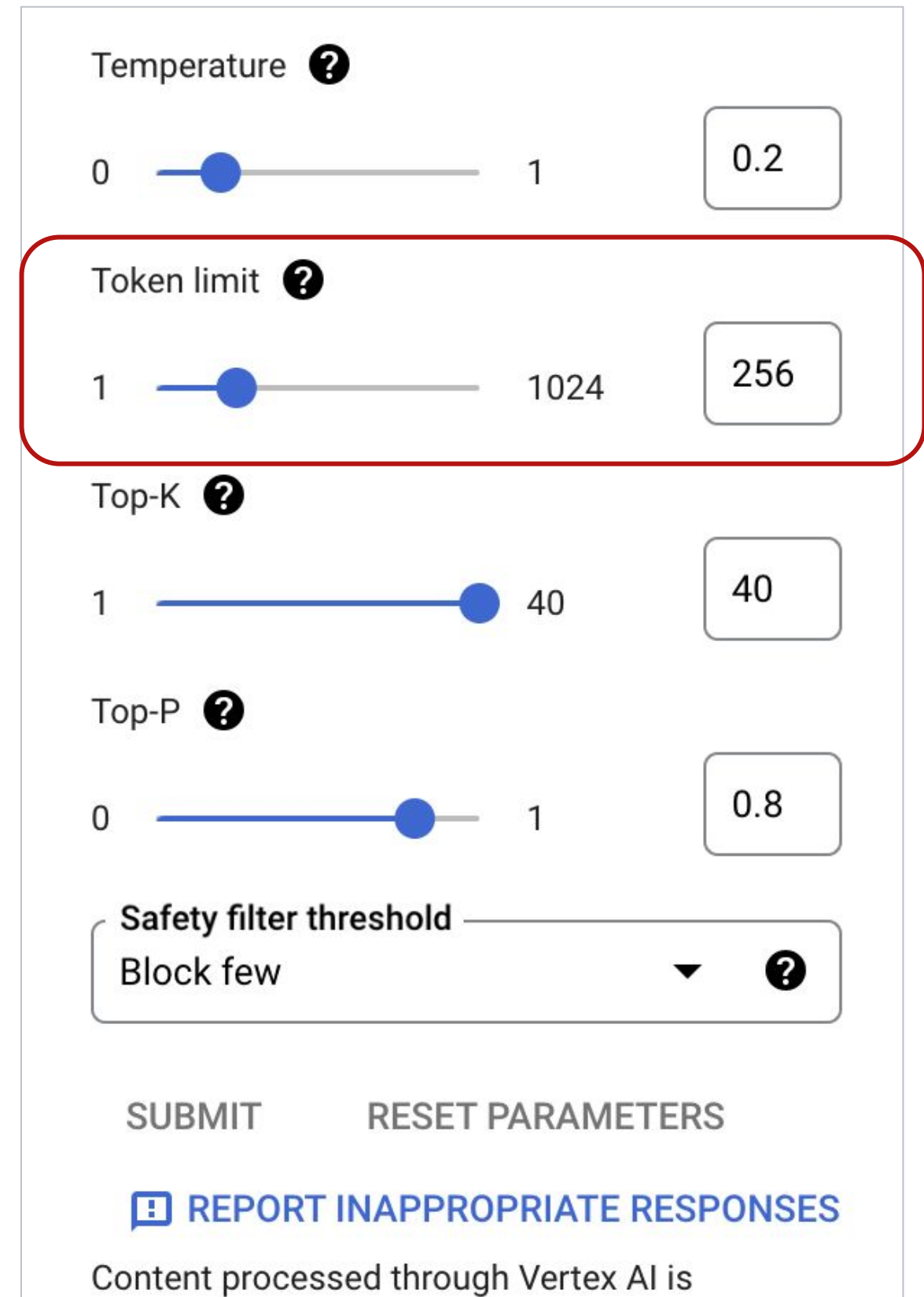
Safety filter threshold: Block few

[REPORT INAPPROPRIATE RESPONSES](#)

Content processed through Vertex AI is

# Token limit determines the maximum amount of output for each prompt

- The range is from 1 to 1024
  - The default value is 256
- A token is a basic unit of text that the LLM understands
  - When using the PaLM API, a token is approximately four characters



Temperature ?

0 1 0.2

Token limit ?

1 1024 256

Top-K ?

1 40 40

Top-P ?

0 1 0.8

Safety filter threshold

Block few ▼ ?

SUBMIT RESET PARAMETERS

[REPORT INAPPROPRIATE RESPONSES](#)

Content processed through Vertex AI is

# Top-K changes how the model selects tokens for output

- Range is from 1 to 40
- A top-K of 1 means the selected token is the most probable among all tokens in the model's vocabulary
- A top-K of 3 means that the next token is selected from among the 3 most probable tokens
- The default top-K value is 40
- The higher the value the more tokens are possible when selecting the next token

Temperature ?

0 —●— 1 0.2

Token limit ?

1 —●— 1024 256

Top-K ?

1 —●— 40 40

Top-P ?

0 —●— 1 0.8

Safety filter threshold ?

Block few ▼

SUBMIT RESET PARAMETERS

[REPORT INAPPROPRIATE RESPONSES](#)

Content processed through Vertex AI is



# Top-P also changes how the model selects tokens for output

- Range is from 0 to 1
- Tokens are selected from most probable to least until the sum of their probabilities equals the top-P value
  - For example, if tokens A, B, and C have a probability of .3, .2, and .1 and the top-P value is .5, then the model will select either A or B as the next token
- The default top-P value is .8
- The higher the value the more likely the model would be to select a token that is not the most probable one

Temperature ?

0 1 0.2

Token limit ?

1 1024 256

Top-K ?

1 40 40

**Top-P ?**

0 1 0.8

Safety filter threshold ?

Block few ▼

SUBMIT RESET PARAMETERS

[REPORT INAPPROPRIATE RESPONSES](#)

Content processed through Vertex AI is

# Topics

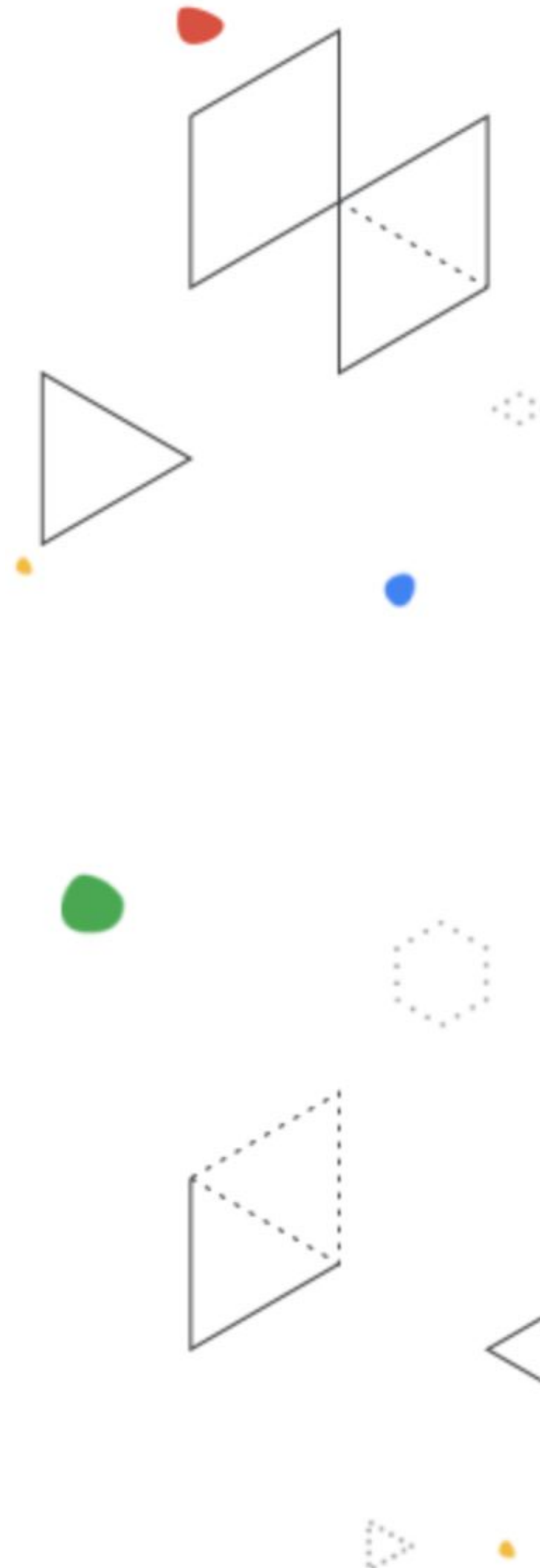
01	Understanding Prompt Design
02	Context, Examples, and Fine-Tuning Parameters
03	Prompt Design Best Practices





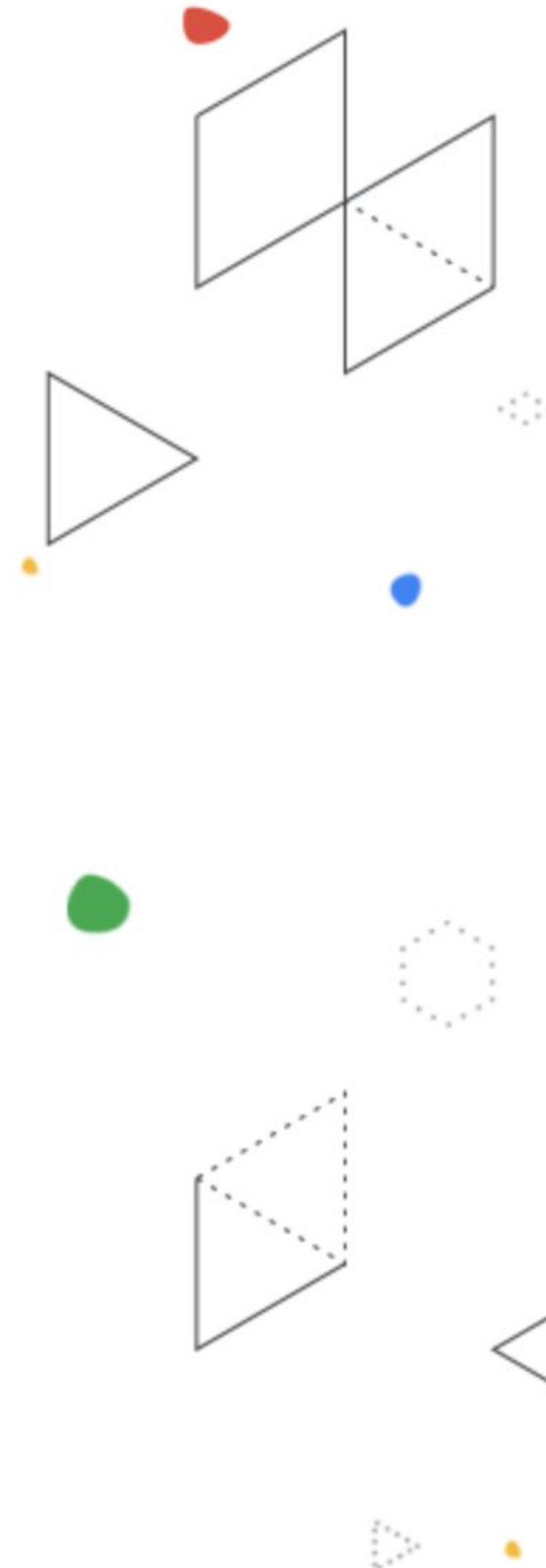
# Prompt design best practices

- Be specific
  - Don't ask: "Write me an a poem"
  - Ask: "Write me a poem about COBOL programming in the style of Edgar Allan Poe"
- Add context
  - Add information about "who" you want the model to emulate
  - What is its job, where does it work, and any other information you would want it to convey
- Add examples
  - Use examples that you have written in the past to teach it to write in your style
  - Give examples of how it should answer questions



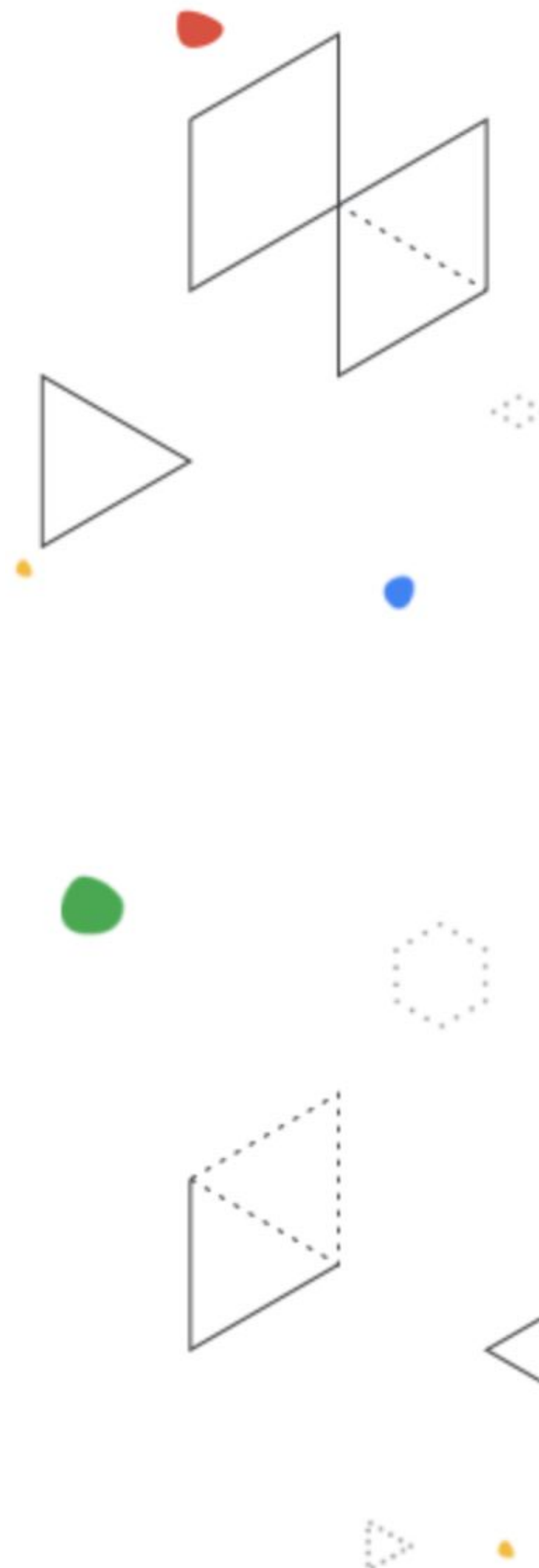
# Be clear and concise

- Use positive language
  - Don't say: "Don't use long sentences"
  - Say: "Write short sentences as bullets"
- Use simple language
  - Avoid jargon
  - Avoid complex sentences
- Use correct grammar and spelling



# Experiment with parameter values

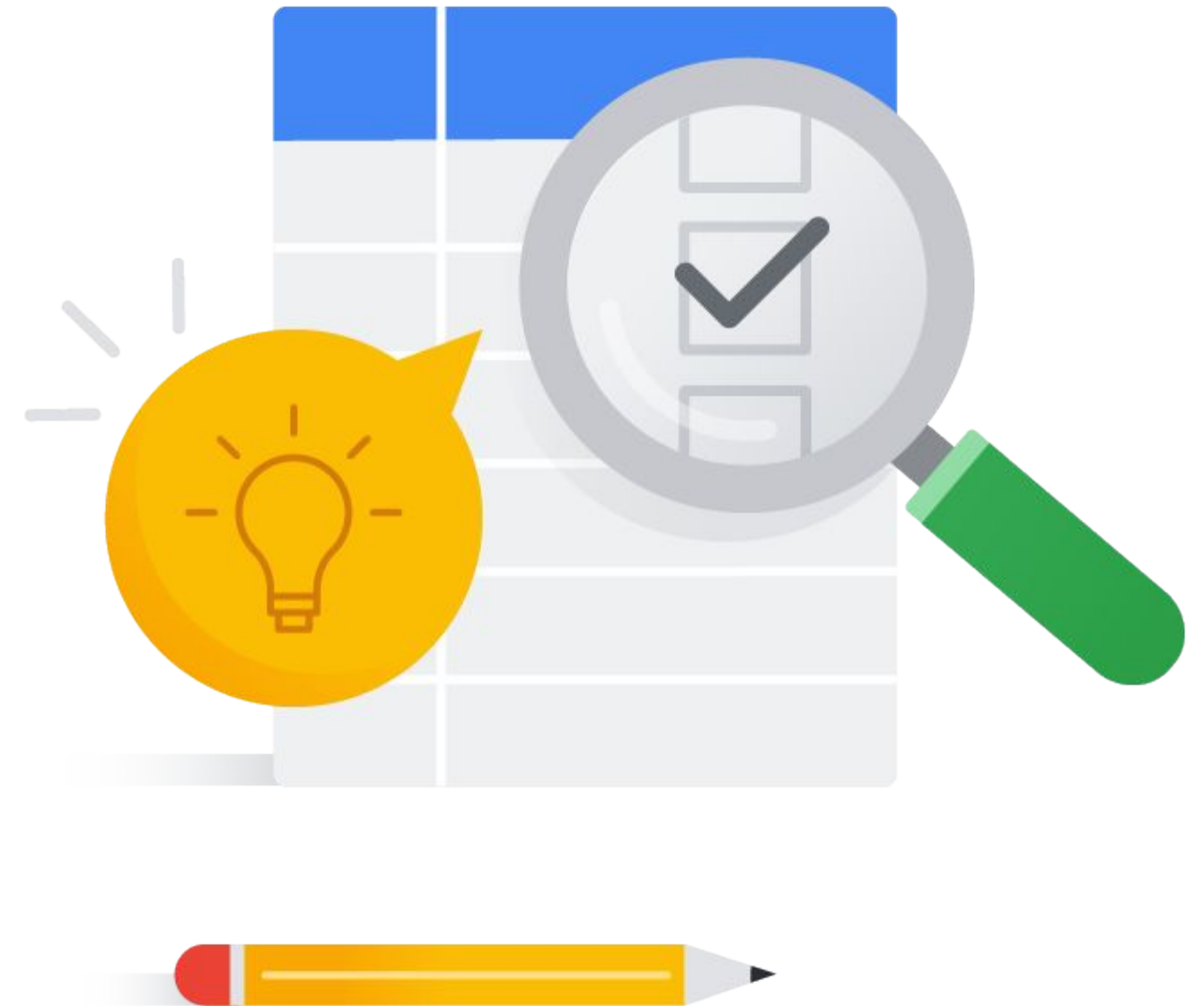
- To make the model more creative, set higher values for Temperature, Top-K, and Top-P
- Set lower values when you want the model to be consistent
- A Temperature of 0, Top-K to 1, and Top-P to 0 would cause the model to always return the same result for a given prompt based on the most likely next token
  - Strictly speaking it is possible to get different results, but it's very rare.



# Lab

🕒 30 min ⚙️

## Prompt Design for Text Chats in Vertex AI Studio

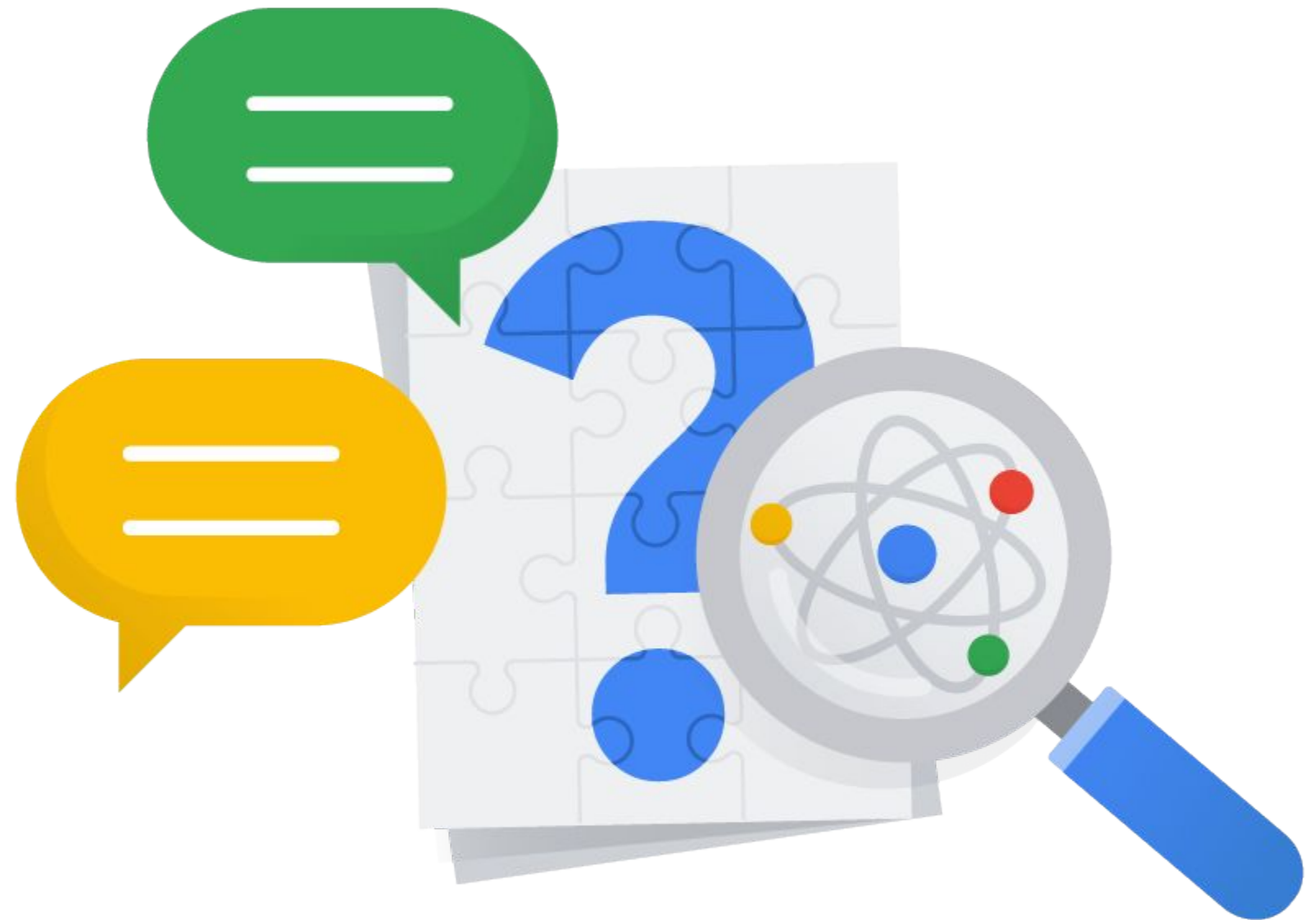


# In this module, you learned to ...

- 01 Design prompts for various Gen AI use cases
- 02 Add examples and context to customize generative AI output
- 03 Experiment with parameters for model fine-tuning



# Questions and answers



# Quiz question

Which setting for temperature would make the model more creative when answering prompts?

A: -1

B: 0

C: 0.1

D: 0.9



# Quiz question

Which setting for temperature would make the model more creative when answering prompts?

A: -1

B: 0

C: 0.1

D: 0.9

# Quiz question

If you want the model to write in your style what should you do?

- A: Set the temperature to about 0.5
- B: Raise the Top-K and Top-P values
- C: Add context
- D: Add examples

# Quiz question

If you want the model to write in your style what should you do?

- A: Set the temperature to about 0.5
- B: Raise the Top-K and Top-P values
- C: Add context
- D: Add examples

# Quiz question

If you want the model to return content more relevant to your organization or use case, what should you do?

- A: Set the temperature to about 0.5
- B: Raise the Top-K and Top-P values
- C: Add context
- D: Add examples

# Quiz question

If you want the model to return content more relevant to your organization or use case, what should you do?

- A: Set the temperature to about 0.5
- B: Raise the Top-K and Top-P values
- C: Add context
- D: Add examples

