

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

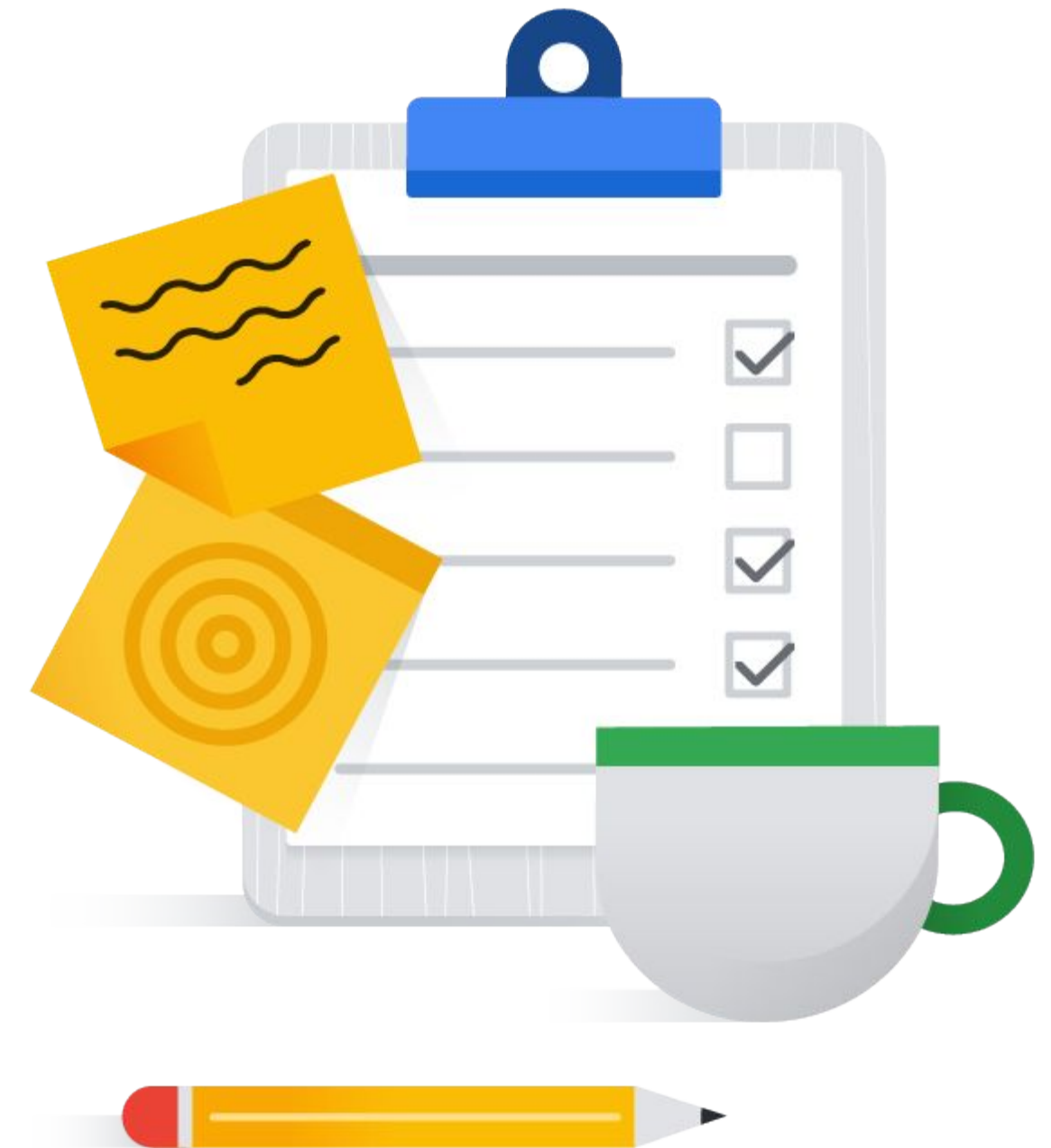
Response

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



Freeform versus Structured prompts in Vertex AI Studio

- So far in class, all the examples shown have been freeform prompts
 - Consist only of the prompt
- Structured prompts in Vertex AI Studio allow you to add context and examples
 - Context tells the model more about your use case characteristics
 - Examples show the model how to structure the response

Vertex AI Studio UI for structured prompts

Examples demonstrate desired output

Context adds information about the use case

Select Structured

Prompt and response

bbq-examples

FREEFORM STRUCTURED

SAVE VIEW CODE DELETE

Context ?

You work for a chef specializing in Southern Barbecue with a large audience on YouTube and social media. You post on Social Media about the latest videos. You like to post with a Southern Style.

Examples ?

INPUT Add an input prefix

OUTPUT Add an output prefix

Write a post about our latest Smoked Brisket video.

Hey y'all, get ready for a flavor-packed ride down South with my latest YouTube video: Smoked Brisket Low and Slow!

Write an input to add a new row

Write a

Test ?

Batch Testing ADD TO EXAMPLES

INPUT Add an input prefix

OUTPUT Add an output prefix

post a review to facebook

Model responses will appear here

Model may display inaccurate or offensive information that doesn't represent Google's view. Not all languages are supported. [Learn more.](#)

We want your [feedback](#).

Region

Temperature ?

0 1 0.2

Advanced

Token limit ?

1 2048 256

Top-K ?

40 40

0.8

1 8 1

Add stop sequence ?

Press Enter after each sequence

Streaming responses ?

Print responses as they're generated

Safety filter threshold ?

Block few

SUBMIT RESET PARAMETERS

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

INPUT	Add an input prefix	OUTPUT	Add an output prefix	<input checked="" type="checkbox"/> Markdown
Write for me an ad for our new orange vest.		<p>Introducing the new orange vest, the perfect way to stay safe and visible on the job.</p> <p>Our orange vest is made from durable, water-resistant material that will keep you protected from the elements. It features reflective strips for added visibility, and a comfortable fit that won't restrict your movement.</p> <p>Whether you're working on a construction site, a farm, or anywhere else where safety is important, our orange vest is the perfect choice for you. Order yours today!</p> <p>Benefits of our orange vest:</p> <ul style="list-style-type: none">• Durable, water-resistant material• Reflective strips for added visibility• Comfortable fit• Perfect for a variety of jobs <p>Order your orange vest today and stay safe on the job!</p>		

Let's see how adding context changes the output

Context ?

You work writing ads in the marketing department for a hunting and fishing supply company. The name of the company is Deep Creek Outfitters. The web site is: www.deepinthecreek.com.

The phone number is: (123) 123-4567

Deep Creek Outfitters

New Orange Vest

Stay Safe and Visible in the Field

Our new orange vest is the perfect way to stay safe and visible in the field. Made from durable, water-resistant material, this vest is perfect for hunting, fishing, or any other outdoor activity. It features a front zipper closure, two front pockets, and a back pocket with a flap. The vest is also adjustable, so you can find the perfect fit.

Order your new orange vest today from Deep Creek Outfitters!

www.deepinthecreek.com

(123) 123-4567

Let's try a different context

✓ Context ?

You write ads for the marketing department of a company called StyleUp that sells clothing targeted at young people in their teens and 20s.

You sell everything online. The website is: www.styleup.cool

Introducing the new orange vest from StyleUp!

This vest is perfect for anyone looking for a stylish and functional piece of outerwear. It's made from a lightweight, water-resistant fabric that will keep you warm and dry on even the coldest days. The orange color is eye-catching and sure to turn heads, and the relaxed fit means you can move freely without feeling restricted.

The vest features two front pockets and a zippered chest pocket, so you have plenty of space to store your belongings. It also has a removable hood that can be zipped on or off depending on the weather.

Whether you're running errands, going to class, or just hanging out with friends, the new orange vest from StyleUp is the perfect way to stay warm and stylish. Order yours today!

Shop now at www.styleup.cool

Let's try requesting a style of output

Context ?

You write ads for the marketing department of a company called StyleUp that sells clothing targeted at young people in their teens and 20s.

You sell everything online. The website is: www.styleup.cool

Your ads are always written in the style of Dr. Seuss.

Oh, the places you'll go In your orange vest!
You'll climb mountains high And swim in the sea.

You'll wear it to school And to the mall. You'll wear it to parties And to the ball.

Your orange vest is so stylish And so cool. It's the perfect way to show Your personal style.

So what are you waiting for? Order your orange vest today! You'll be glad you did.

Visit www.styleup.cool to order yours now!

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

Context

You work for a chef specializing in Southern Barbecue with a large audience on YouTube and social media. You post on Social Media about the latest videos. You like to post with a Southern Style.

Examples

IMPORTADD INPUT COLUMN

INPUT	OUTPUT
Add an input prefix	Add an output prefix
Write a post about our latest Smoked Brisket video.	Get ready for a flavor-packed ride down South with my latest YouTube video: Smoked Brisket Low and Slow! Slow-smoked to juicy perfection, seasoned with secret spices. Mouthwatering sides included! Join me on the grill, y'all! #SouthernBBQ #SmokedBrisket #BBQPassion
Write an input to add a new row	Write an output to add a new row

Test

Batch TestingADD TO EXAMPLES

INPUT	OUTPUT
Add an input prefix	Add an output prefix
Write a post about our new video about smoking Prime Rib for the holidays.	Y'all better sit down for this one! My new video is all about how to smoke a Prime Rib for the holidays. It's gonna be a real showstopper, packed with flavor and juicy goodness. Get ready to impress your family and friends with this holiday classic. #PrimeRib #HolidayCooking #BBQ #SouthernStyle

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

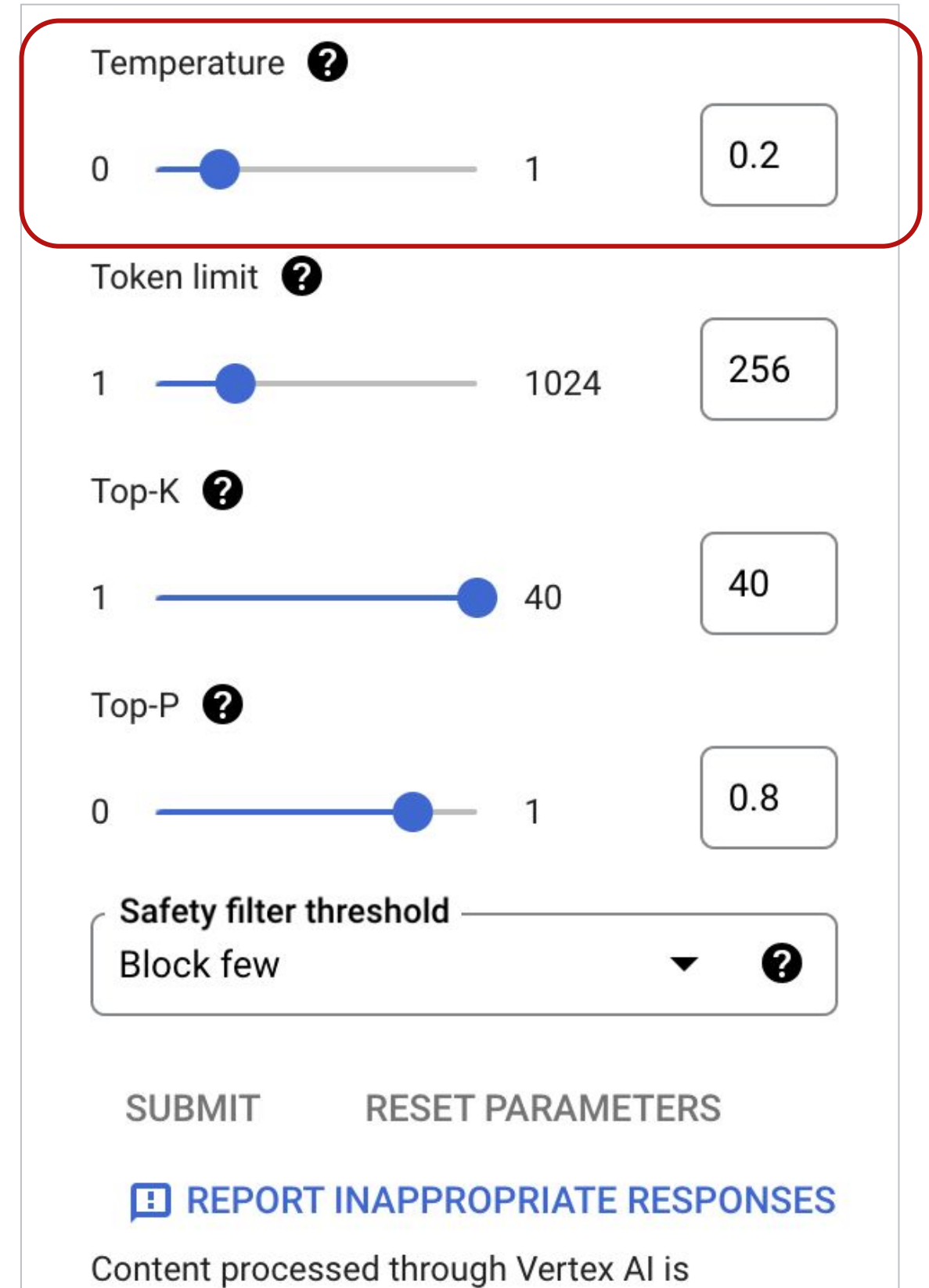
[SUBMIT](#) [RESET PARAMETERS](#)

[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 control panel for AI generation 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

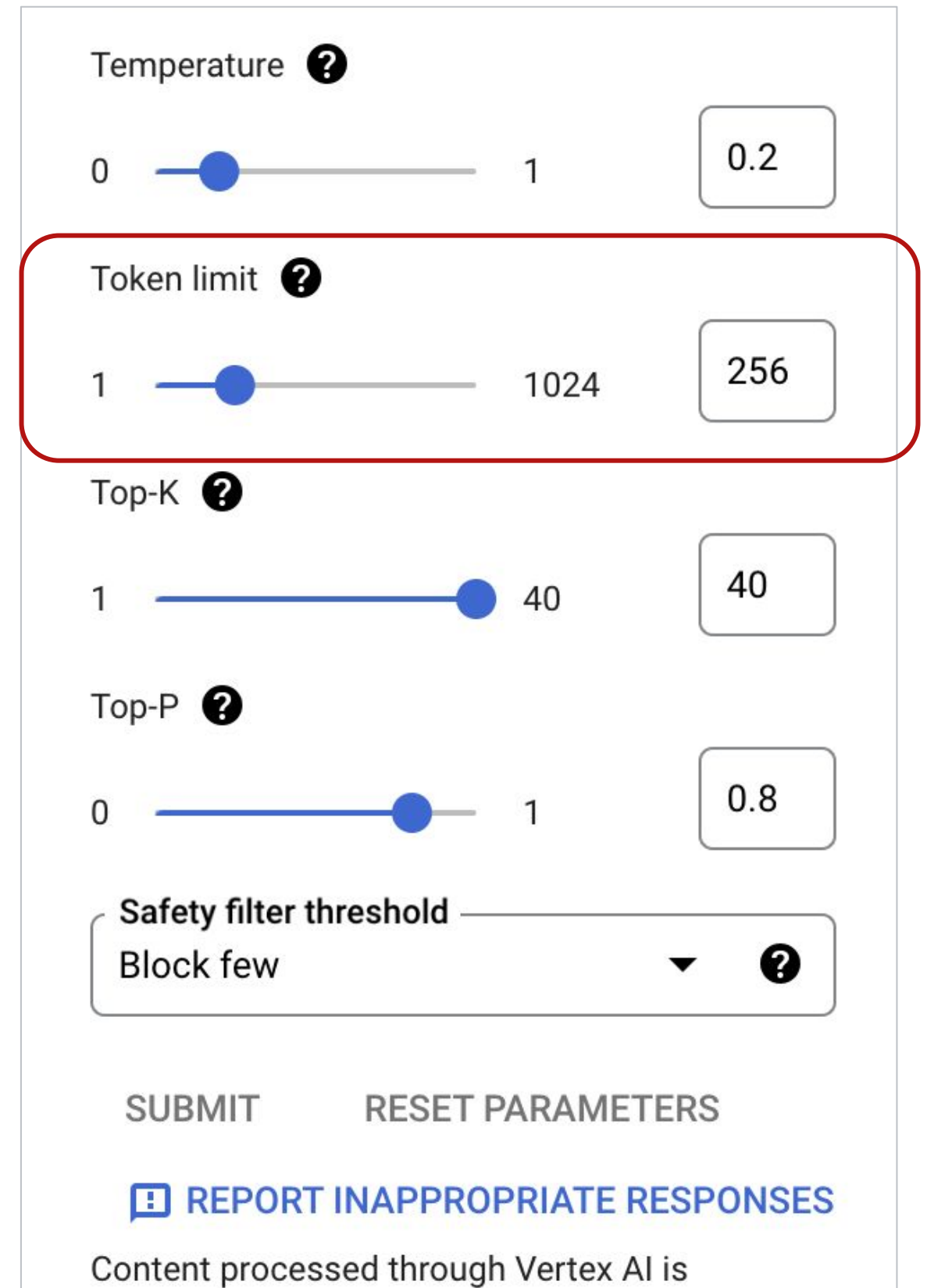
Safety filter threshold: Block few

[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
- A token is a string of characters that represents a concept or object in GenAI
 - For example, the token "dog" represents the concept of a dog



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
 - Called greedy decoding
 - See: https://www.tensorflow.org/text/guide/decoding_api
- 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

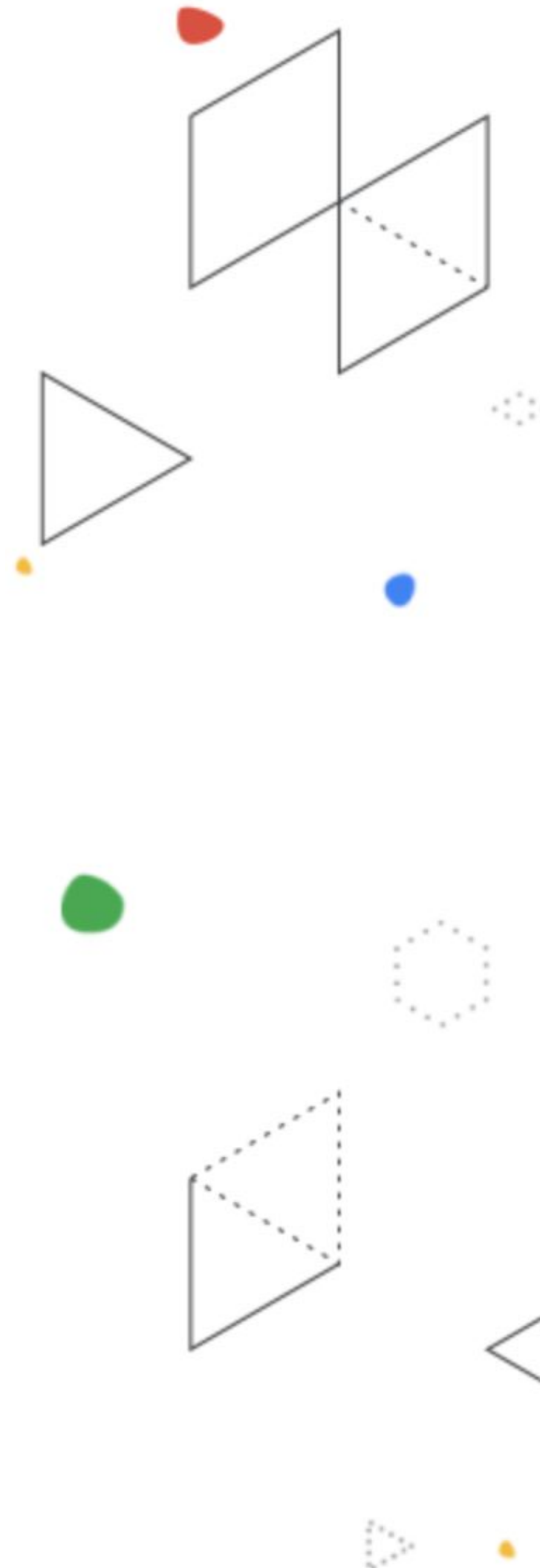
Topics

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



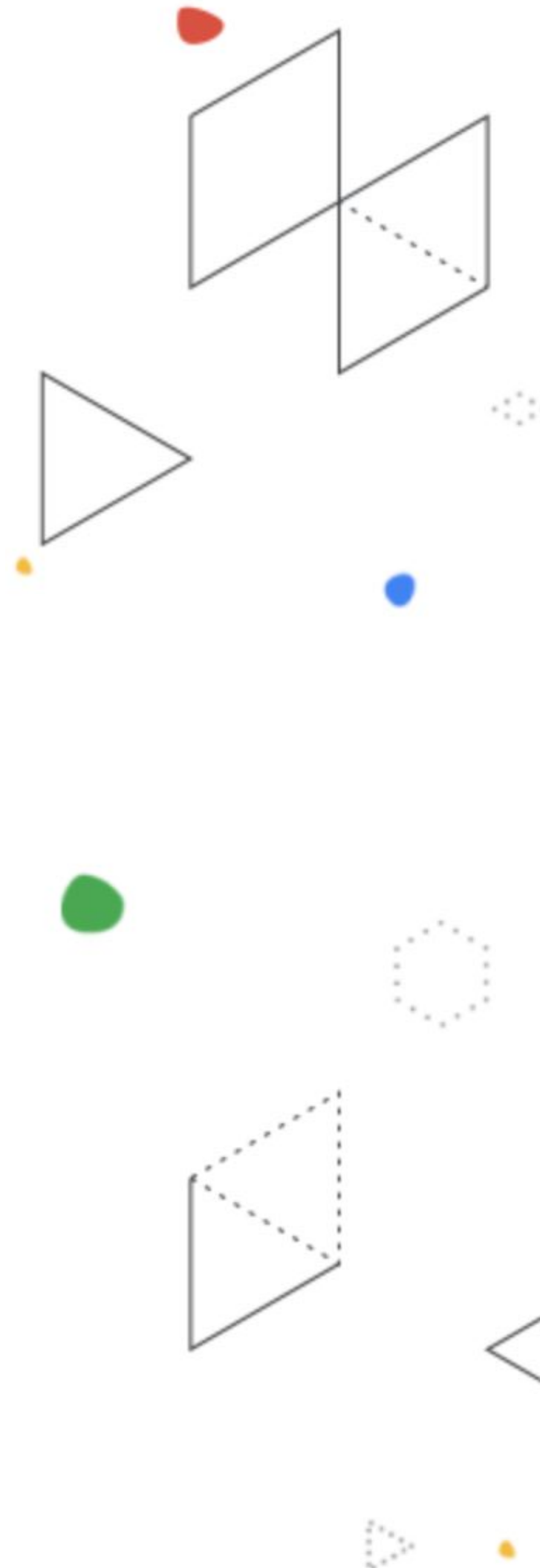
Prompt design best practices

- 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



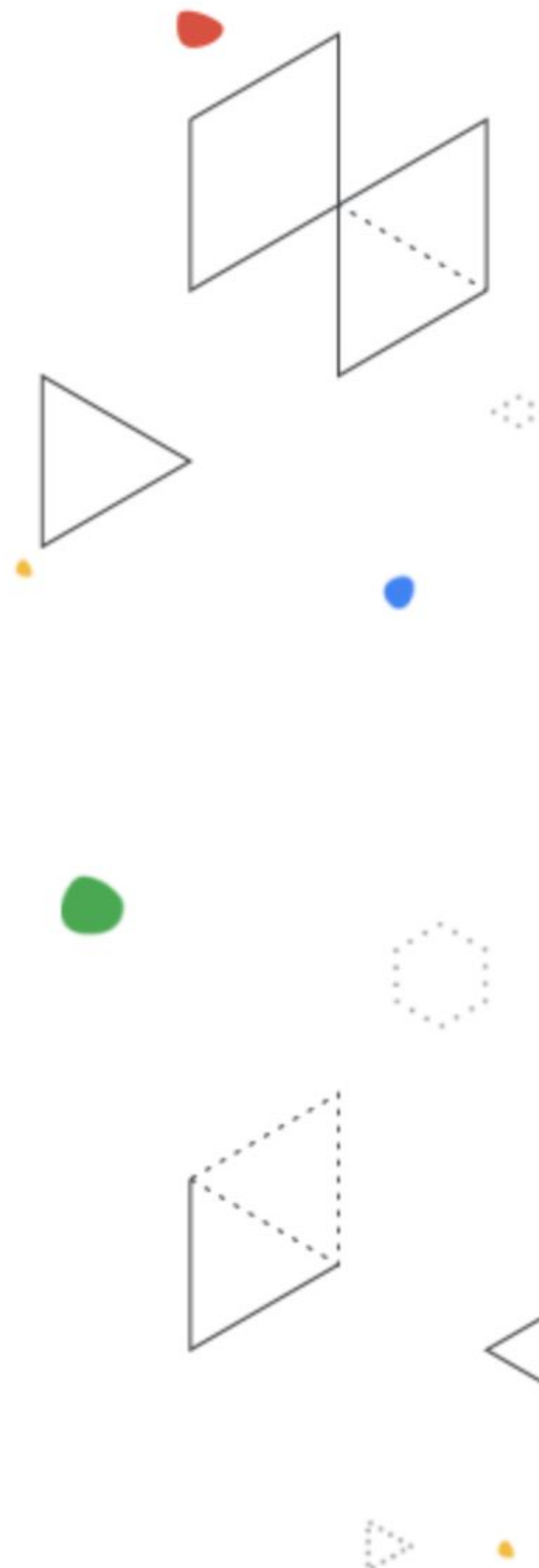
Be clear and concise

- 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

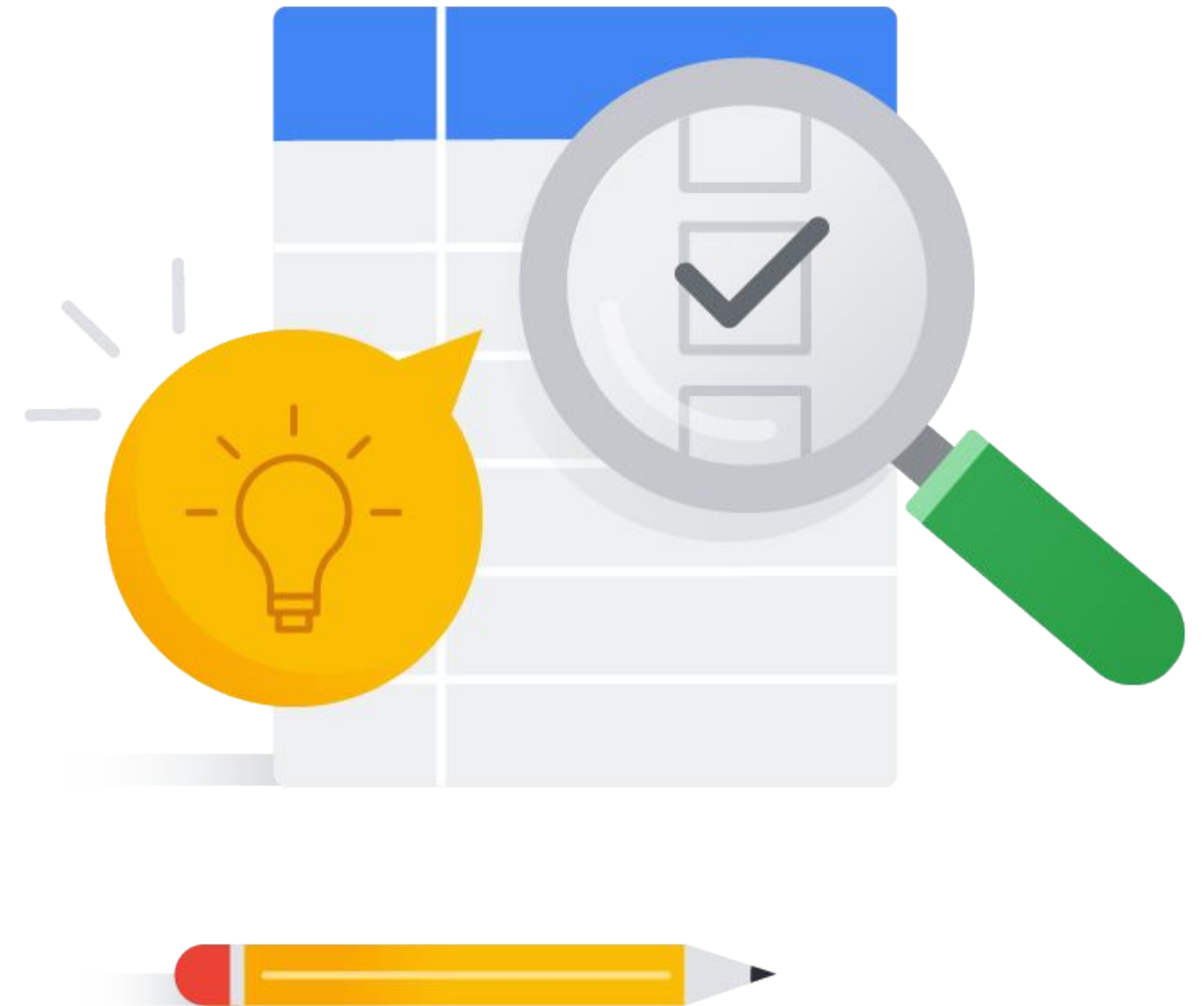
- To make the model more creative, set higher values for Temperature, Top-K, and Top-P
- Set lower values when you want the model to be consistent
- A Temperature of 0, Top-K to 1, and Top-P to 0 would cause the model to always return the same result for a given prompt based on the most likely next token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



Lab

🕒 30 min ⚙️

Prompt Design for Text Generation in Vertex AI Studio



In this module, you learned to ...

01

Design prompts for various Gen AI use cases

02

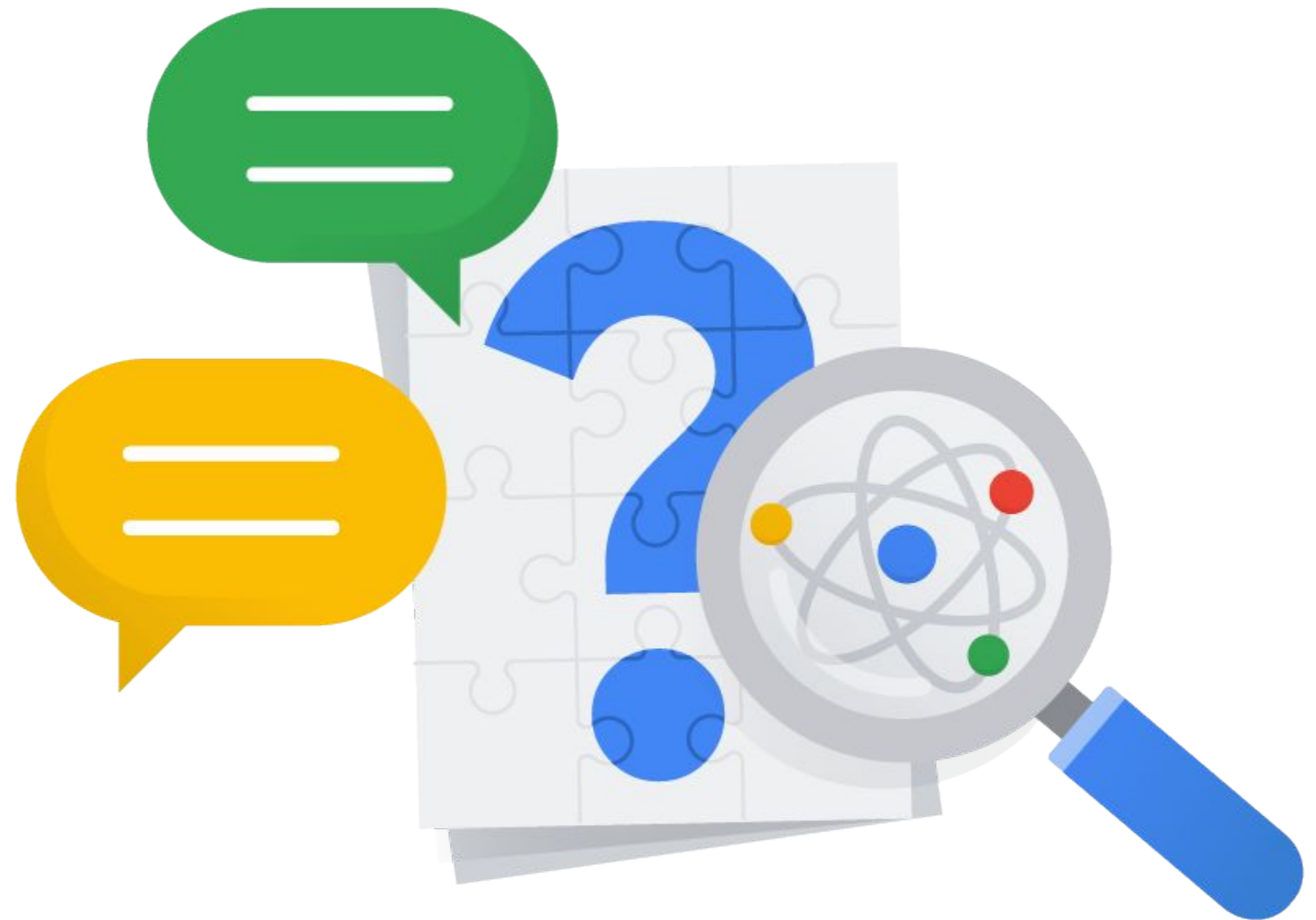
Add examples and context to customize generative AI output

03

Experiment with parameters for model fine-tuning



Questions and answers



Quiz question

What's the difference between freeform and structured prompts?

Quiz question

What's the difference between freeform and structured prompts?

With structured prompts you can add context and examples.

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

