Responsible AI

INTRODUCTION TO GPTS

Jasmin Ludolf

Data Science and AI Content Developer, DataCamp

Responsible AI

> The practice of developing and using AI systems in a way that benefits society while minimizing the risk of negative consequences

[1] https://www.iso.org/artificial-intelligence/responsible-ai-ethics

Responsible AI

> The practice of developing and **using** AI systems in a way that benefits society while minimizing the risk of negative consequences

- **Everyone** plays a part in safe AI use
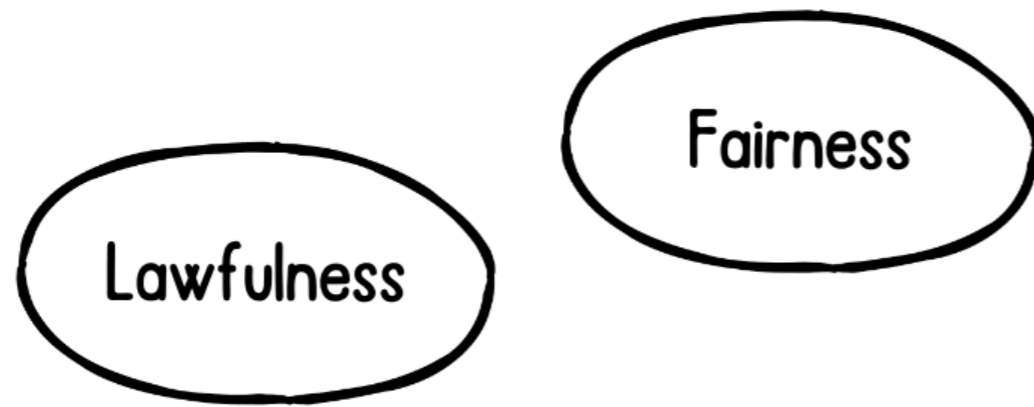
- We can't blame the tool!

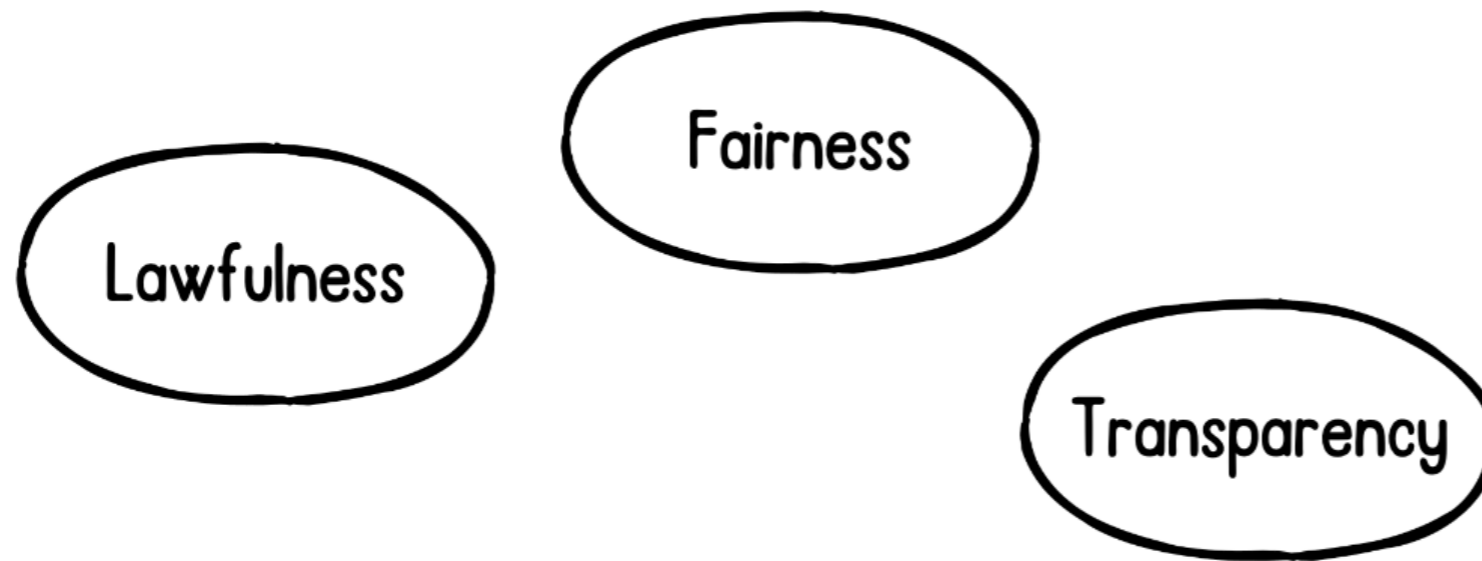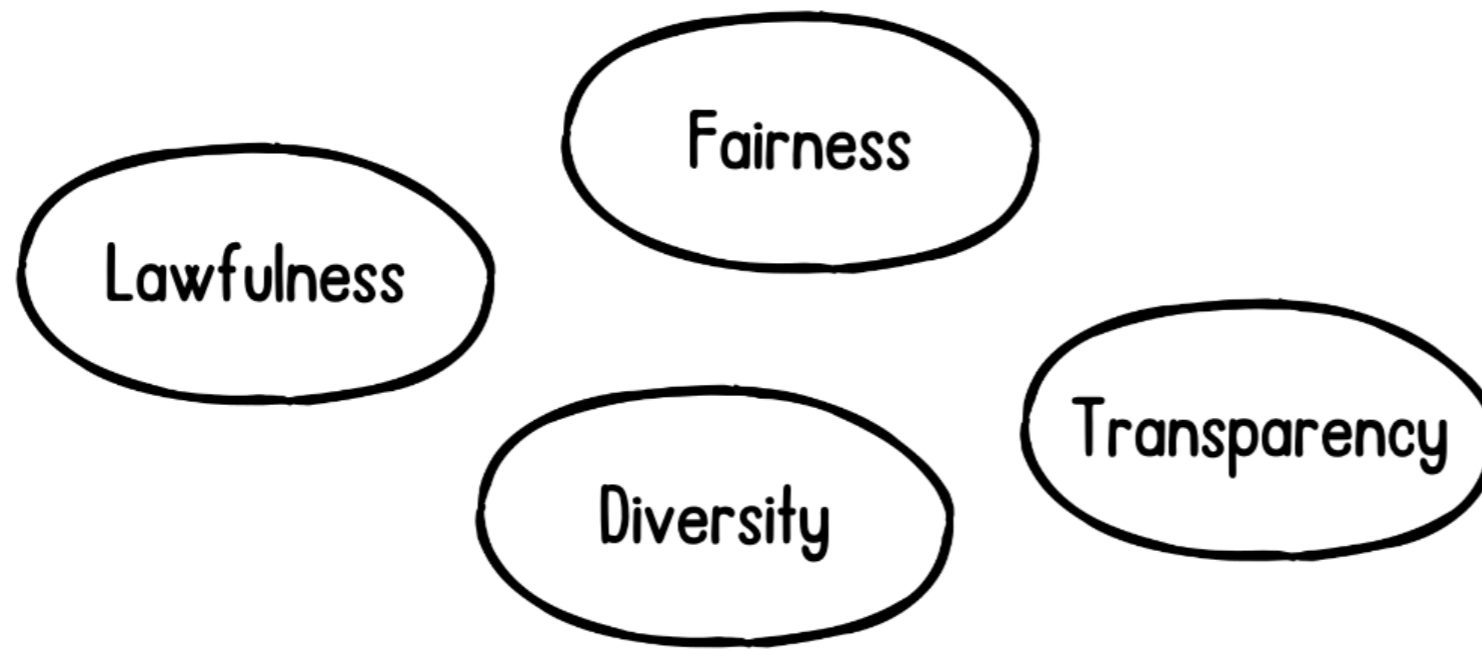[1] https://www.bbc.co.uk/news/world-us-canada-65735769

Considerations

Lawfulness

Considerations

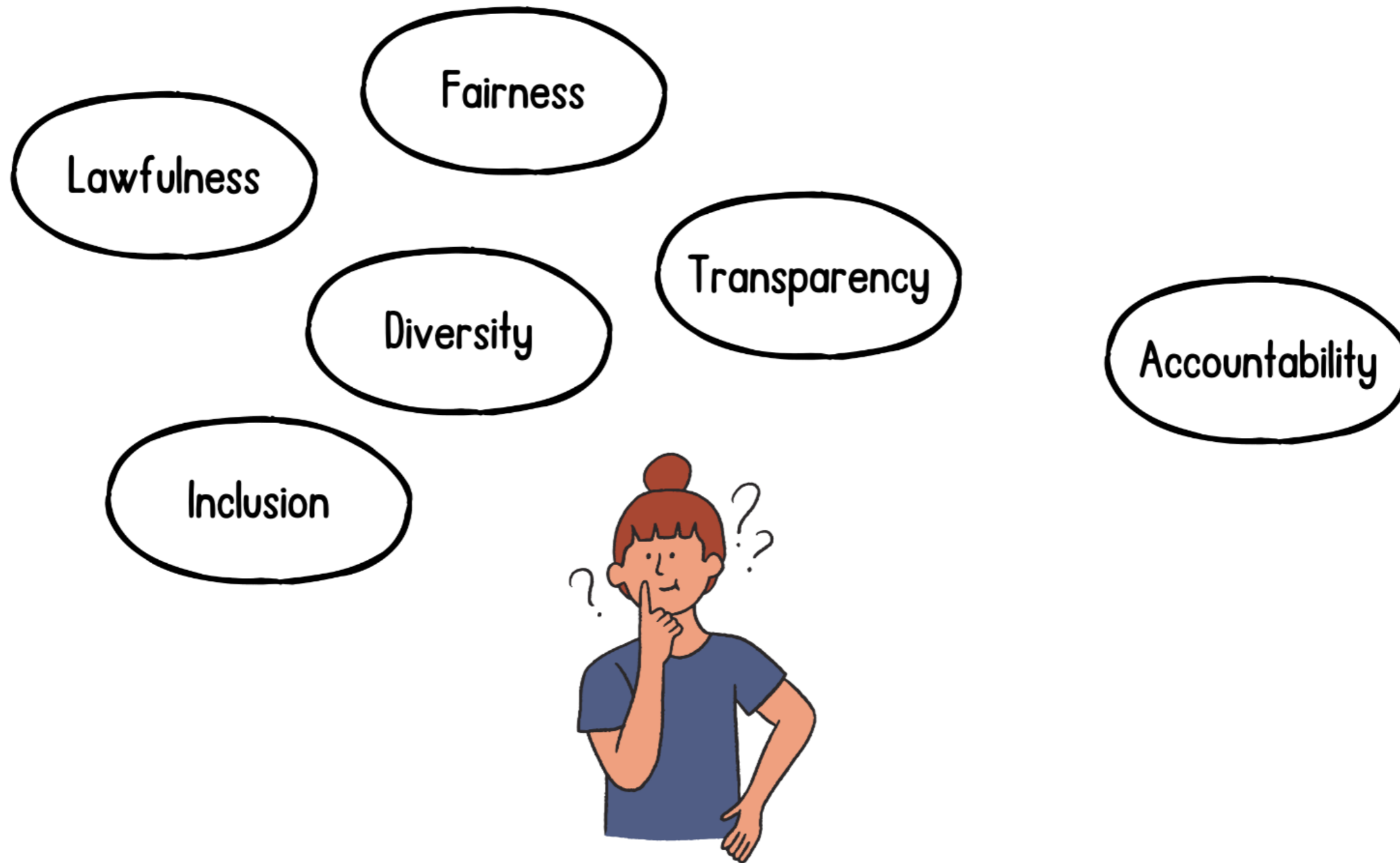Lawfulness

Fairness

Considerations

Lawfulness

Fairness

Transparency

Considerations

Fairness

Lawfulness

Transparency

Diversity

Considerations

Considerations

Considerations

Fairness

Lawfulness

Transparency

Diversity

Accountability

Inclusion

Privacy

Considerations

Fairness

Security

Lawfulness

Transparency

Diversity

Accountability

Inclusion

Privacy
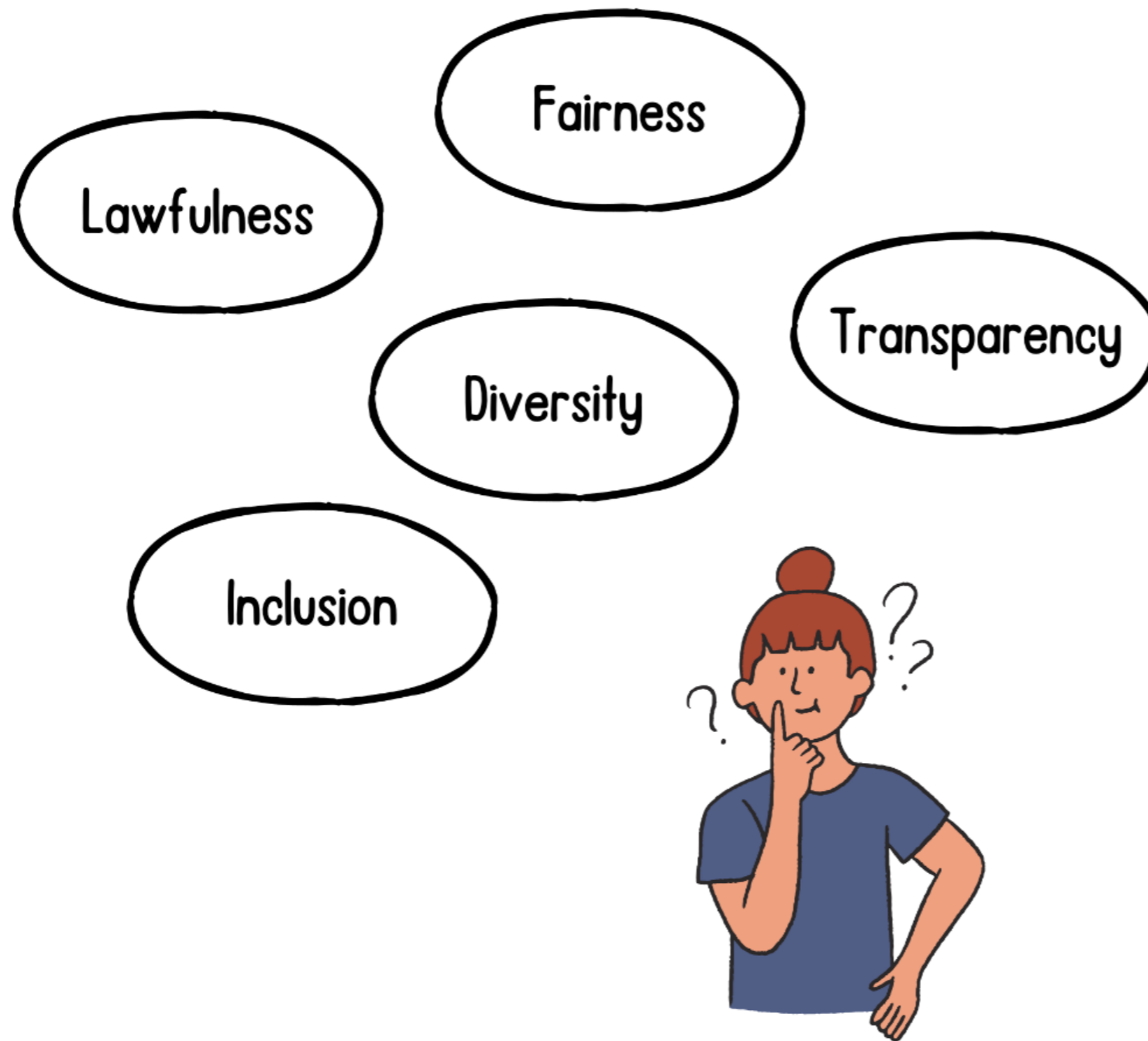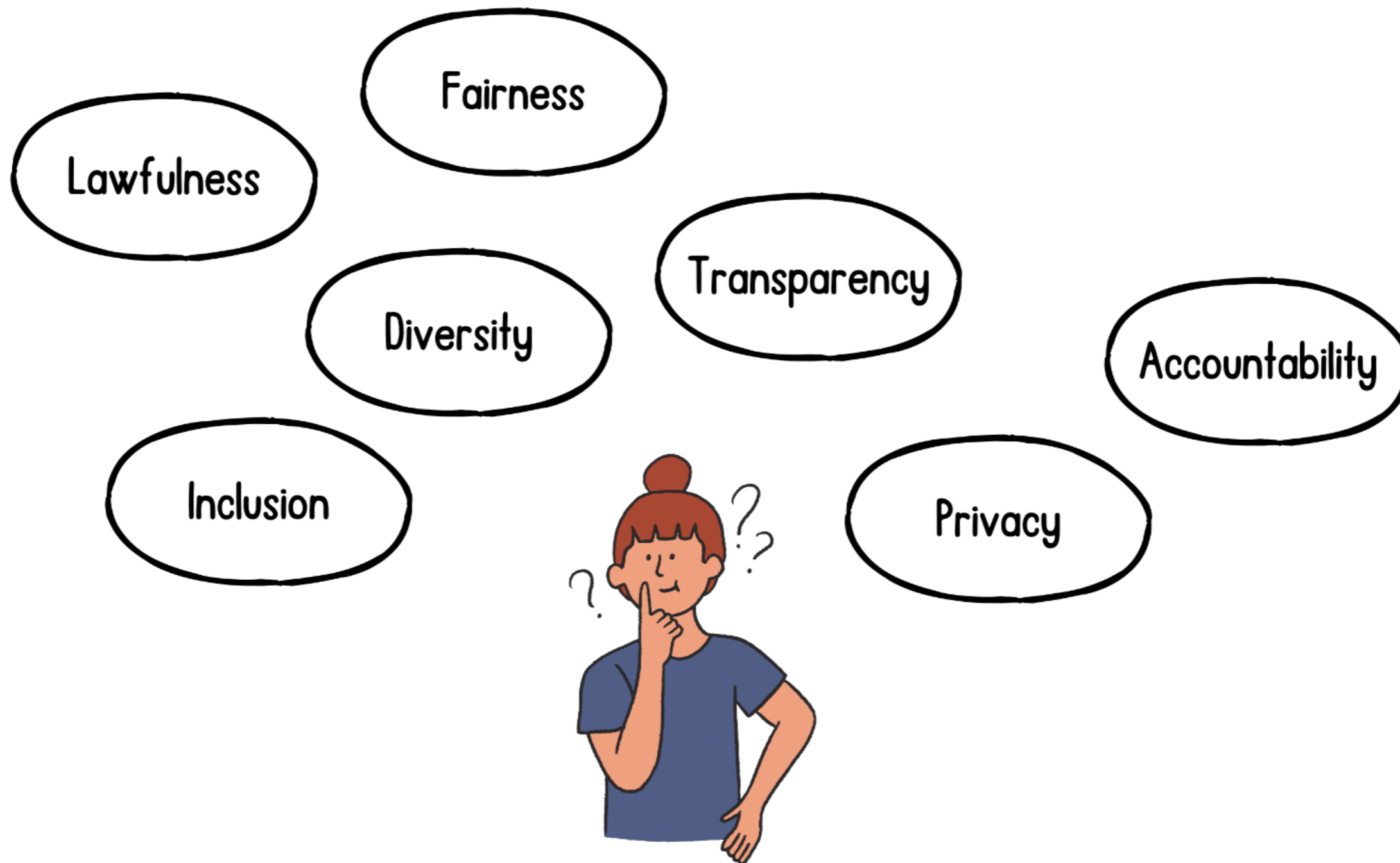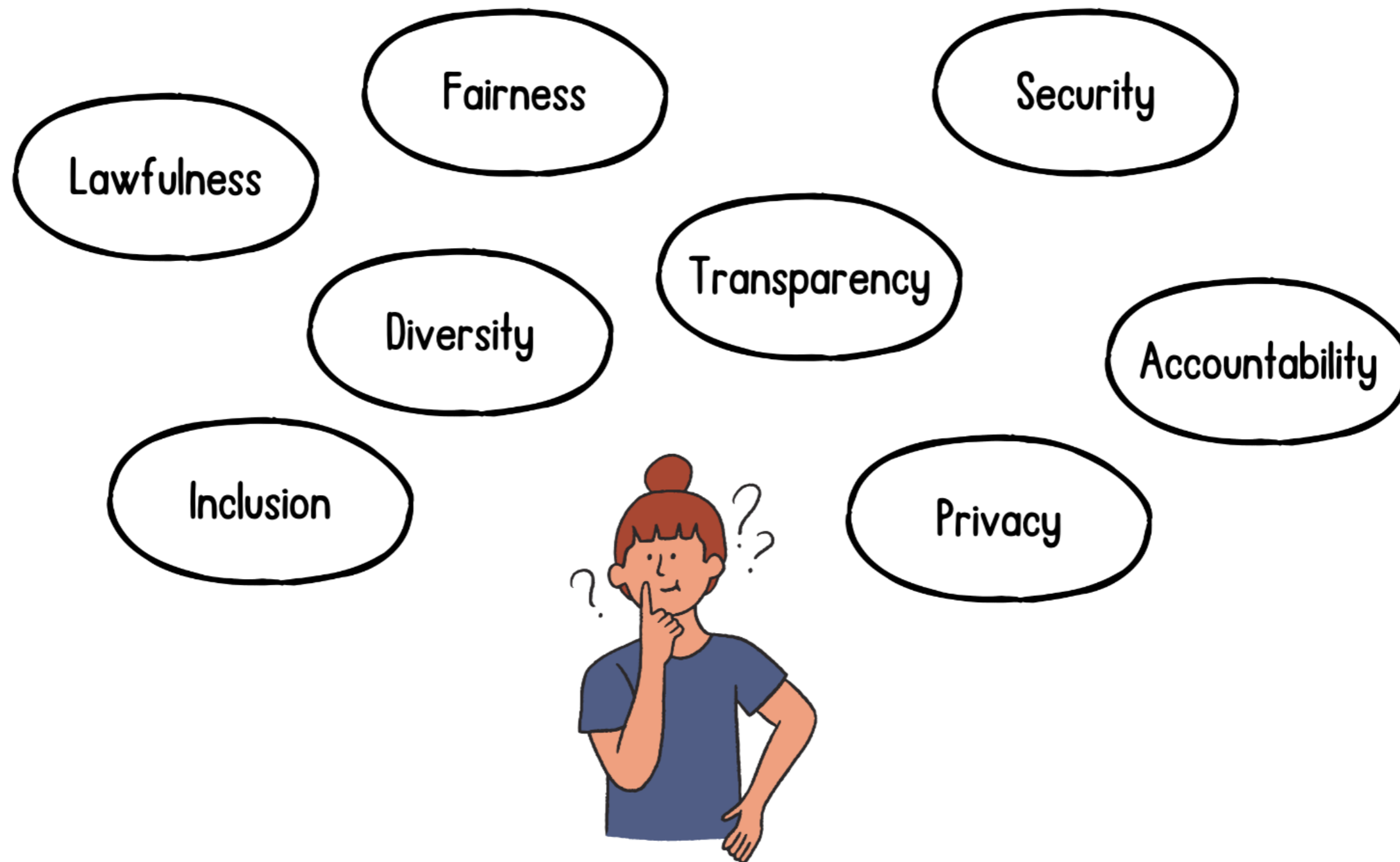
User considerations

- Behavior complies with laws and regulations
  - Different depending on location and industry

- Treat everyone equally without discrimination

- Understand how the tool works

- Respect individual rights

- Protect business data

- Applies to prompts and outputs

Check the policy

- Code of conduct or policy

- How to use it

- Data expectations

- ⬜ Follow these policies

Guidelines on:

- Financial data

- Personal data

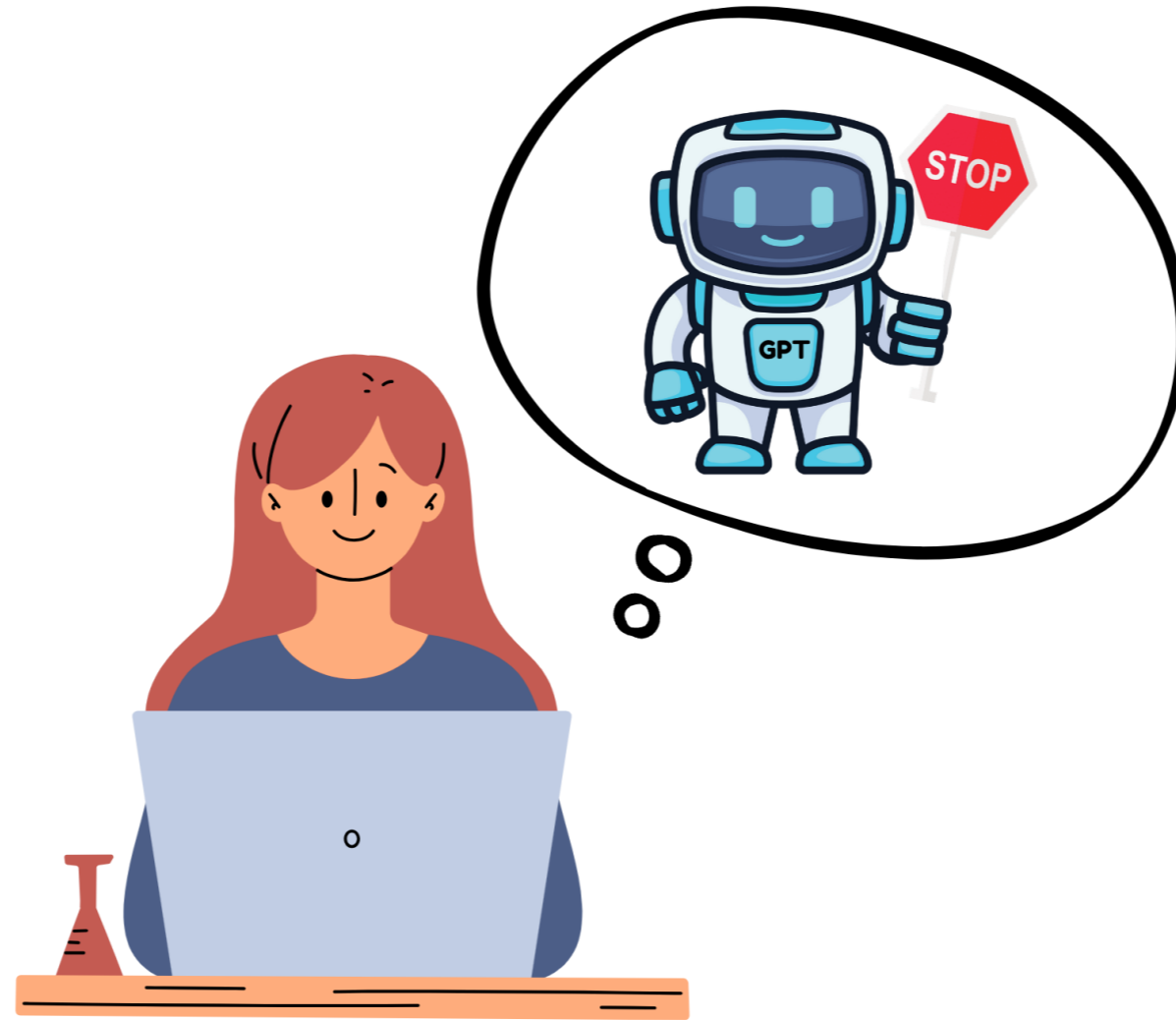- Proprietary data

Policies may differ!

General input considerations

- Personally identifiable information:
  - ID numbers

  - Addresses

  - Medical

  - Banking

- Proprietary information:
  - Confidential strategies

  - Legal

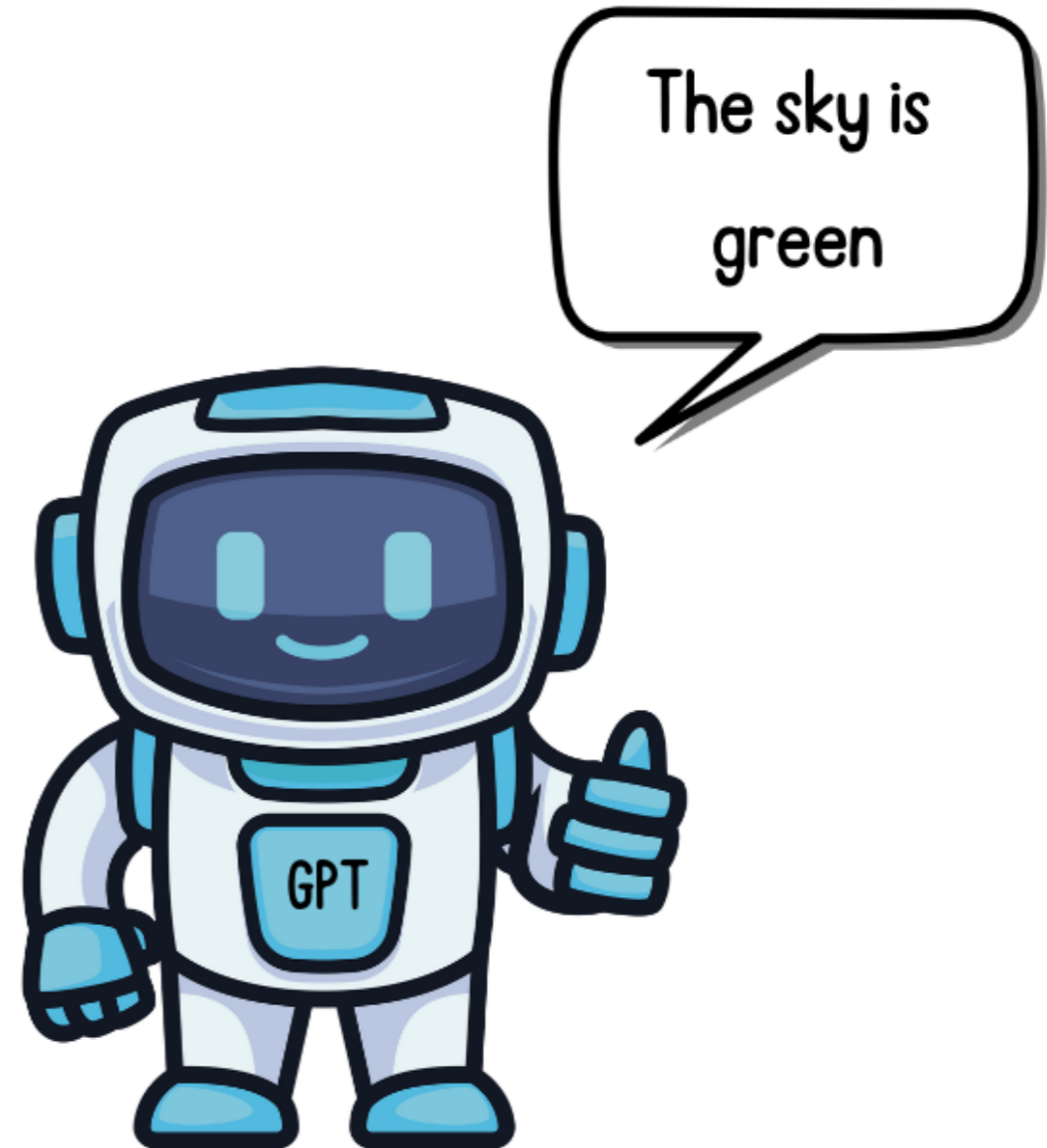 - Financial

General input considerations

- Provider using the data

- Vulnerable to attacks

General output considerations

- Bias: unfairly favors one outcome

- Context tracking: losing track of the conversation

- Hallucination: confident incorrect answer

No 100% guarantee

Our responsibility

The sky is green

Let's practice!

INTRODUCTION TO GPTS

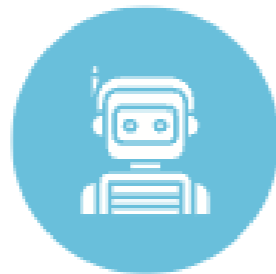# GPT inputs

INTRODUCTION TO GPTS

Jasmin Ludolf

Data Science and AI Content Developer,
DataCamp

Prompt engineering

Can you help me with this task?

How can I help?

Prompt interpretation

Write a product description for our new shirt.

Prompt interpretation

Write a **product description** for our **new** shirt.

Prompt interpretation

Write a product description for our new shirt.

Techniques

Specific
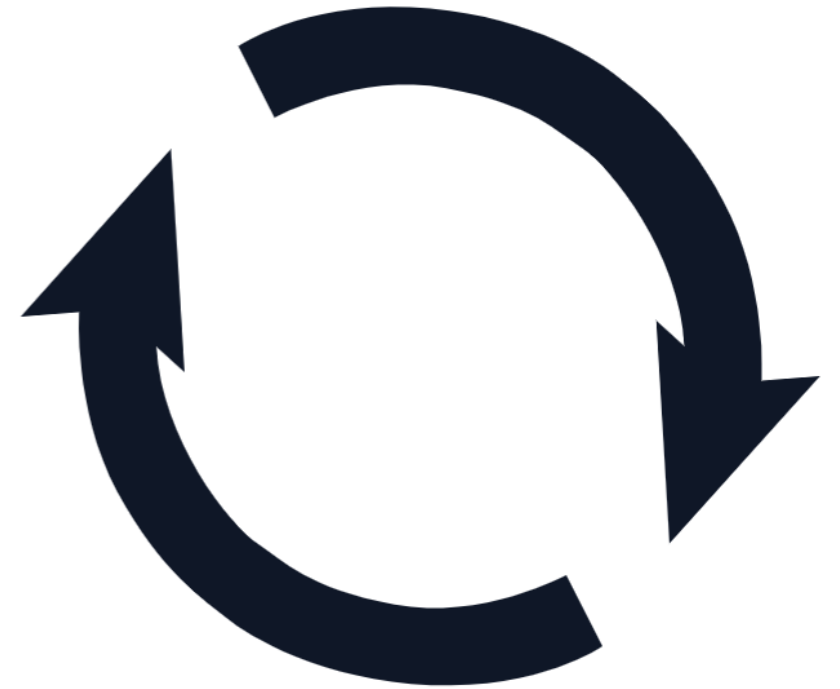
Techniques

Specific

Examples

Techniques

Specific

Examples

Iterate

Be specific

Write a product description for our shirt

Be specific

Write a product description for our new blue linen ID 45 shirt.

Be specific

Write a product description for our new blue linen ID 45 shirt. Please describe the shirt, including the cut and color, with bullet points.

Be specific

Write a product description for our new blue linen ID 45 shirt. Please describe the shirt, including the cut and color, with bullet points. Include one sentence that summarizes the full description.

Provide an example

Write a product description for our new blue linen ID 45 shirt. Please describe the shirt, including the cut and color, with bullet points. Include one sentence that summarizes the full description. Follow this example:

Product title: Red Graphic T-Shirt
Description: This red graphic t-shirt offers soft cotton comfort and a bold design, perfect for casual wear.
   ○ Short sleeve
   ○ Round neck
   ○ 100% cotton

Provide a template

Write a product description for our new blue linen ID 45 shirt. Please describe the shirt, including the cut and color, with bullet points. Include one sentence that summarizes the full description. Follow this example:

Product title: Red Graphic T-Shirt
Description: This red graphic t-shirt offers soft cotton comfort and a bold design, perfect for casual wear.
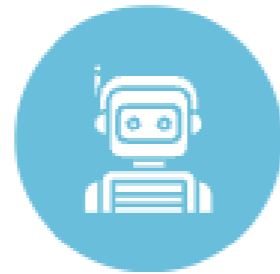 ○ Short sleeve
  ○ Round neck
  ○ 100% cotton

📎 Fill in this template

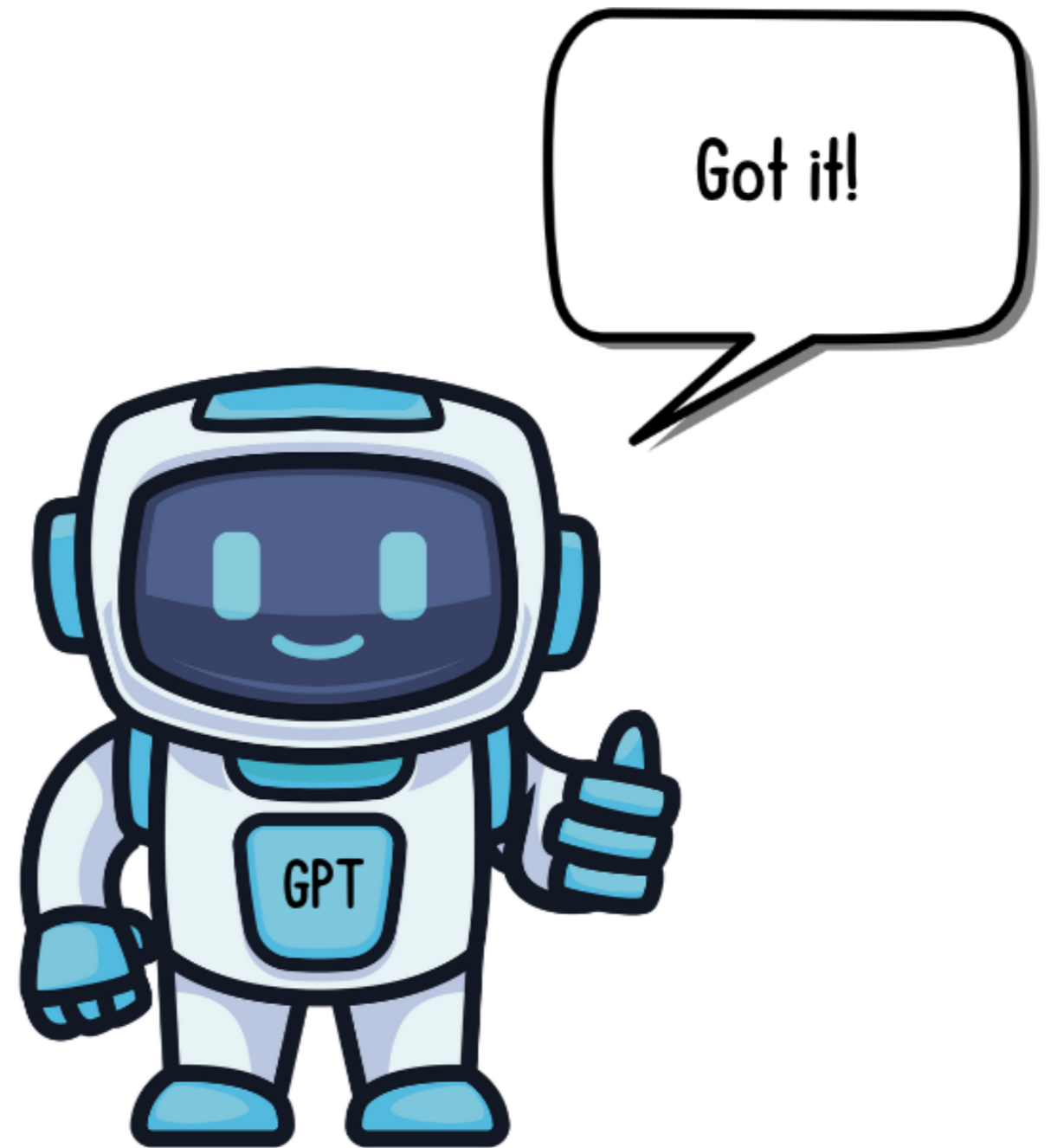# Iterate

Confidential tasks

- Tool will have built-in safeguards

- Recommended to vet prompt

- Confirm prompts are legally sound **before** they are put into the tool

[1] https://legalbriefs.deloitte.com/post/102j486/prompt-engineering-for-legal-security-and-risk-considerations

Effective prompts

- Be specific

- Context

- Instructions

- Examples

- Iterate

- Vetting
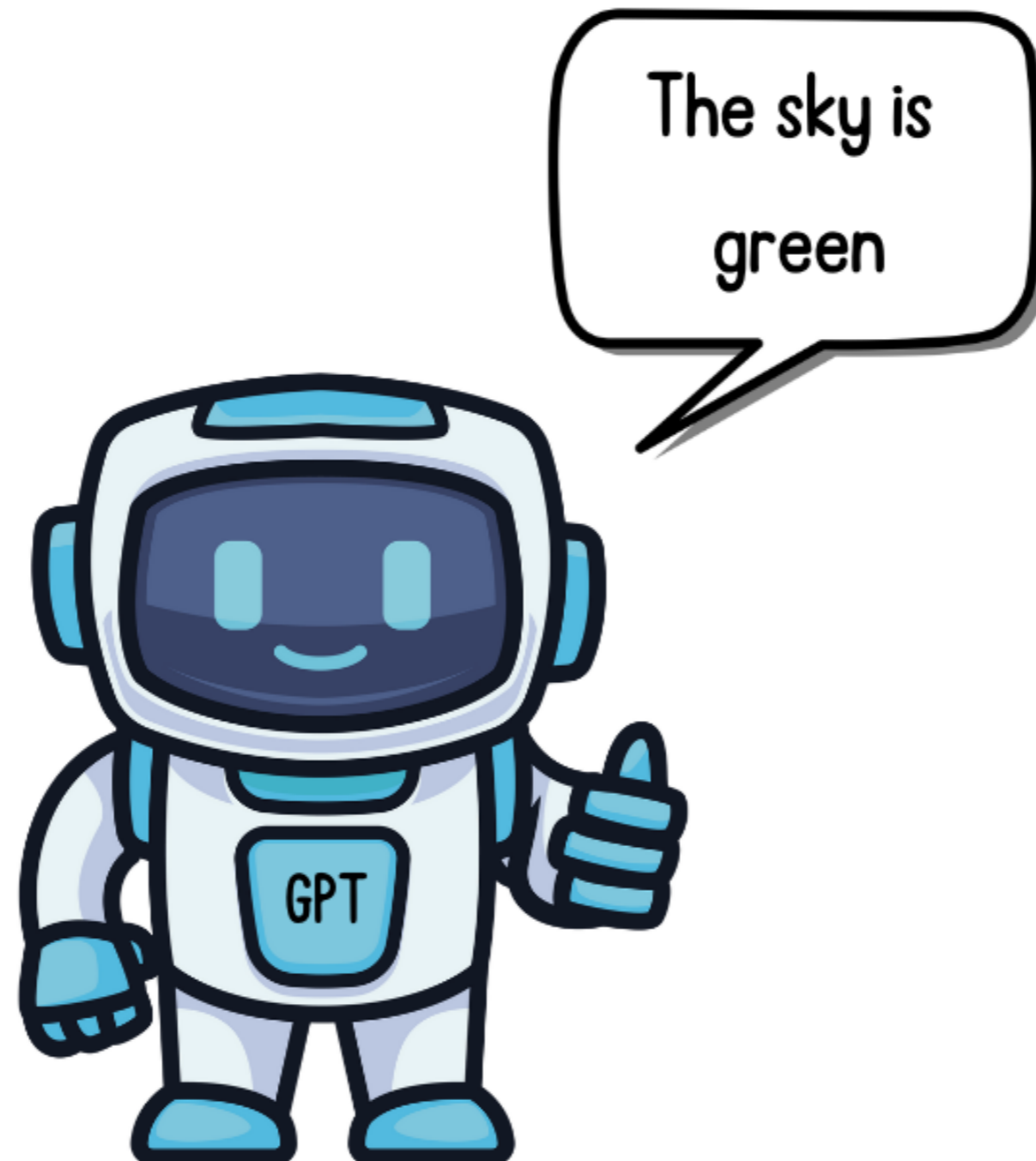
Let's practice!

datacamp

GPT outputs

INTRODUCTION TO GPTS

Jasmin Ludolf

Data Science and AI Content Developer,
DataCamp

Incorrect outputs

Validation techniques

- Evaluate the output ourselves with domain knowledge

- Compare the results or ask an expert

- Use question-and-answering

- Always proofread

Validate ourselves

- Financial analyst:
  - Transfers a route calculation

  - Checks it with their own calculation

Comparing results

- Analyst writing a report:
  - Compare the results with a previous report
  - Compare with other company sources

Ask an expert

- People team manager:
  - Creating a week-long program

  - Connects with benefits manager

  - Connects with mental health specialist

  - Checks the plan is appropriate

Question-and-answering

- Stock manager:
  - Ask follow-up questions to verify the answer
  - "What was your methodology?"
  - "Show me the steps you took to get to this answer"
  - "Please correct this part"

Let's practice!

INTRODUCTION TO GPTS