



CS329X: Human Centered NLP

Human-AI Collaboration

Diyi Yang

Stanford CS

Overview

Overview of human-model interaction

Key desiderata for human-model collaboration

“Collaborating with LLMs” –prompting

Slides credit to Sherry Wu

Human-AI Interaction: What is it?

Basically, a field where **humans** and **AIs** **interact**.

Humans: AI researchers, model developers, domain experts, end users.

AIs: dialog system, translator, recommender system, autonomous driving system.

Interact:

Humans collaborate with AI,

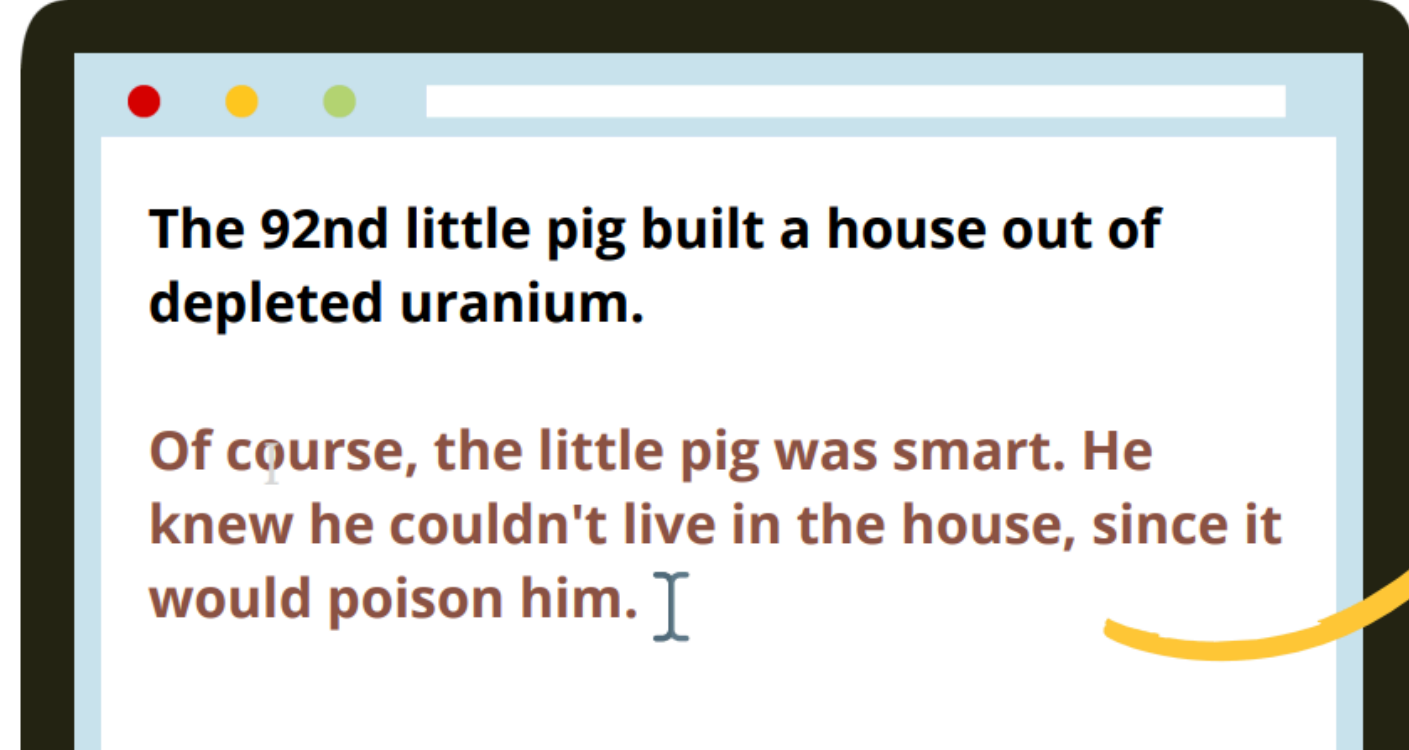
Humans get assistance from AI-infused applications,

Humans analyze AI

Human-AI Collaboration

The **cooperative** and **coordinated** interaction between humans (mostly non-AI experts) and AI to solve *complex problems* or *achieve certain goals*.

Write

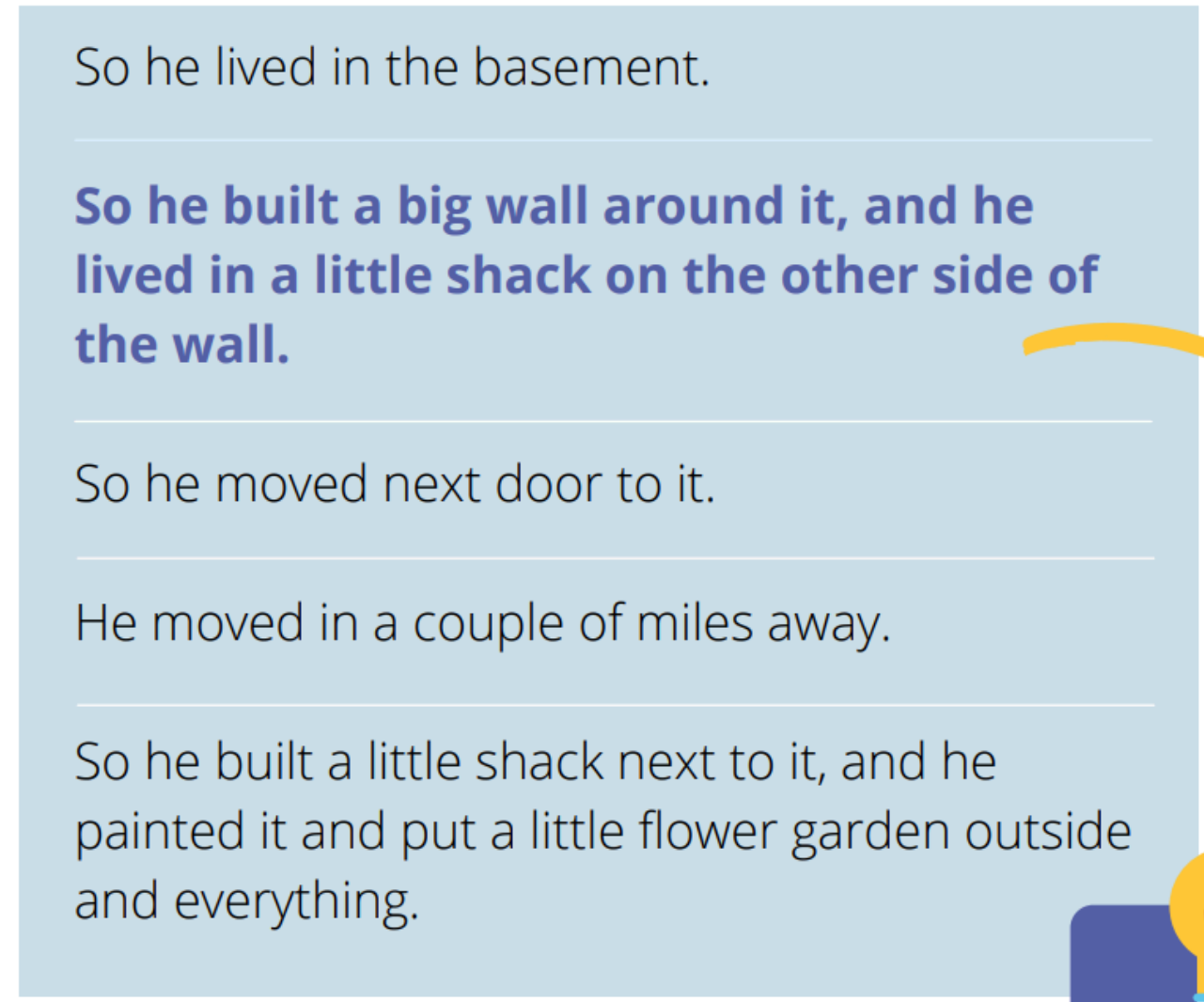


The 92nd little pig built a house out of depleted uranium.

Of course, the little pig was smart. He knew he couldn't live in the house, since it would poison him. I



Get suggestions



So he lived in the basement.

So he built a big wall around it, and he lived in a little shack on the other side of the wall.

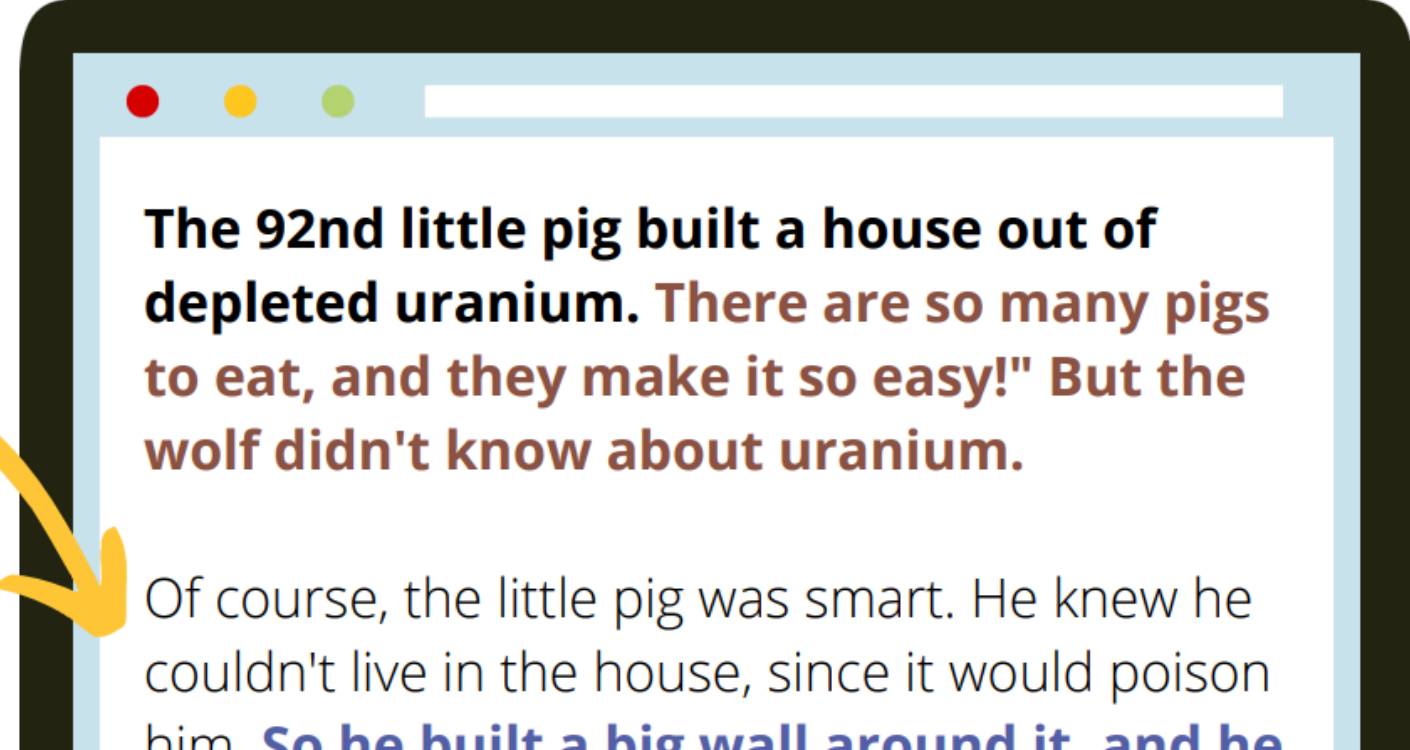
So he moved next door to it.

He moved in a couple of miles away.

So he built a little shack next to it, and he painted it and put a little flower garden outside and everything.



Edit



The 92nd little pig built a house out of depleted uranium. **There are so many pigs to eat, and they make it so easy!" But the wolf didn't know about uranium.**

Of course, the little pig was smart. He knew he couldn't live in the house, since it would poison him. **So he built a big wall around it, and he lived in a little, hidden shack on the other side of the wall.**



Humans get assistance from AI-infused applications

Similar to humans interacting with models: humans are still mostly **end users** and **domain experts**. The big difference is **AI is not a partner**, but a tool (and part of "AI-infused applications")

Because we want people to get **smooth assistance** from AIs when they are in the **larger application context** (e.g., Amazon suggestion page is only one section), **the concept of task & AI model is blurred**.

because these models are **wrapped under mature visual interfaces**, people tend to **have less tolerance when they get wrong**.



Humans analyze Models

So AI experts can systematically understand ML models, and go beyond aggregated scores.

“Understanding the **broader terrain** of errors is an important starting point in pursuing systems that are robust, safe, and fair... [We need to] identify **cohorts** with higher error rates and diagnose the **root causes** behind these errors.”

Eric Horvitz / Microsoft, 2021

<https://erroranalysis.ai/>

Human-AI Interaction: What is it?

Basically, a field where **humans** and **AIs** **interact**.

Humans: AI researchers, model developers, domain experts, end users.

AIs: dialog system, translator, recommender system, autonomous driving system.

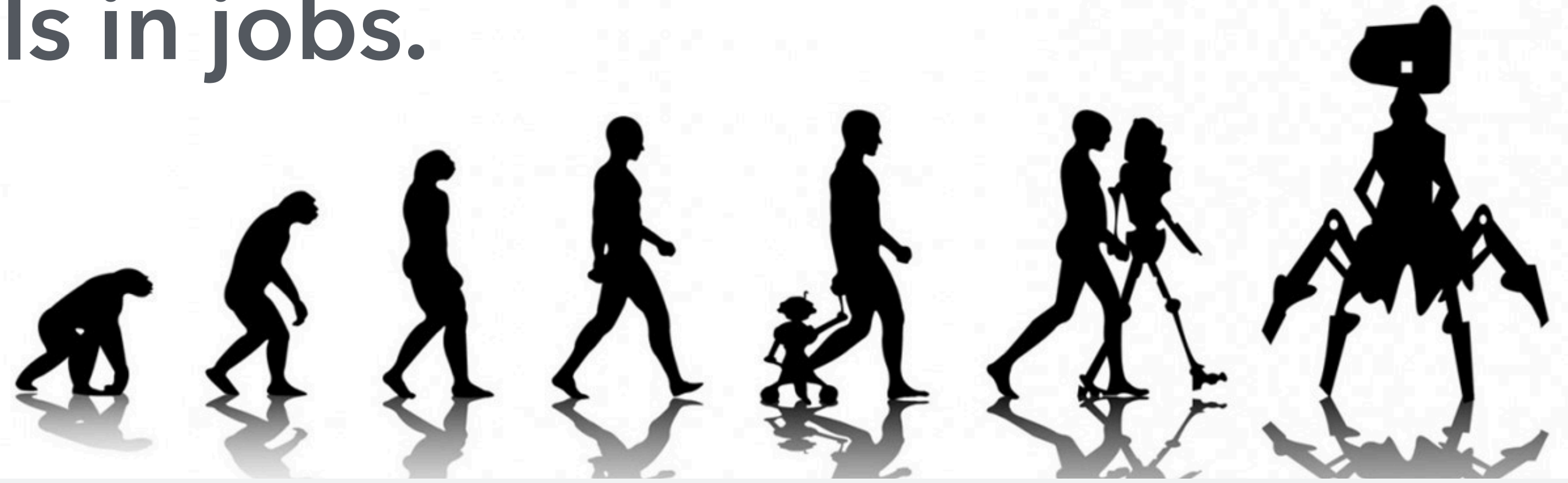
Interact:

Humans collaborate with AI,

Humans get assistance from AI-infused applications,

Humans analyze AI

AI does not automate jobs. It automates certain skills in jobs.



AI will indeed automate most repetitive and physical tasks...and will push human professionals up the skillset ladder into uniquely human skills such as creativity, social abilities, empathy, and sense-making, which machines cannot automate.

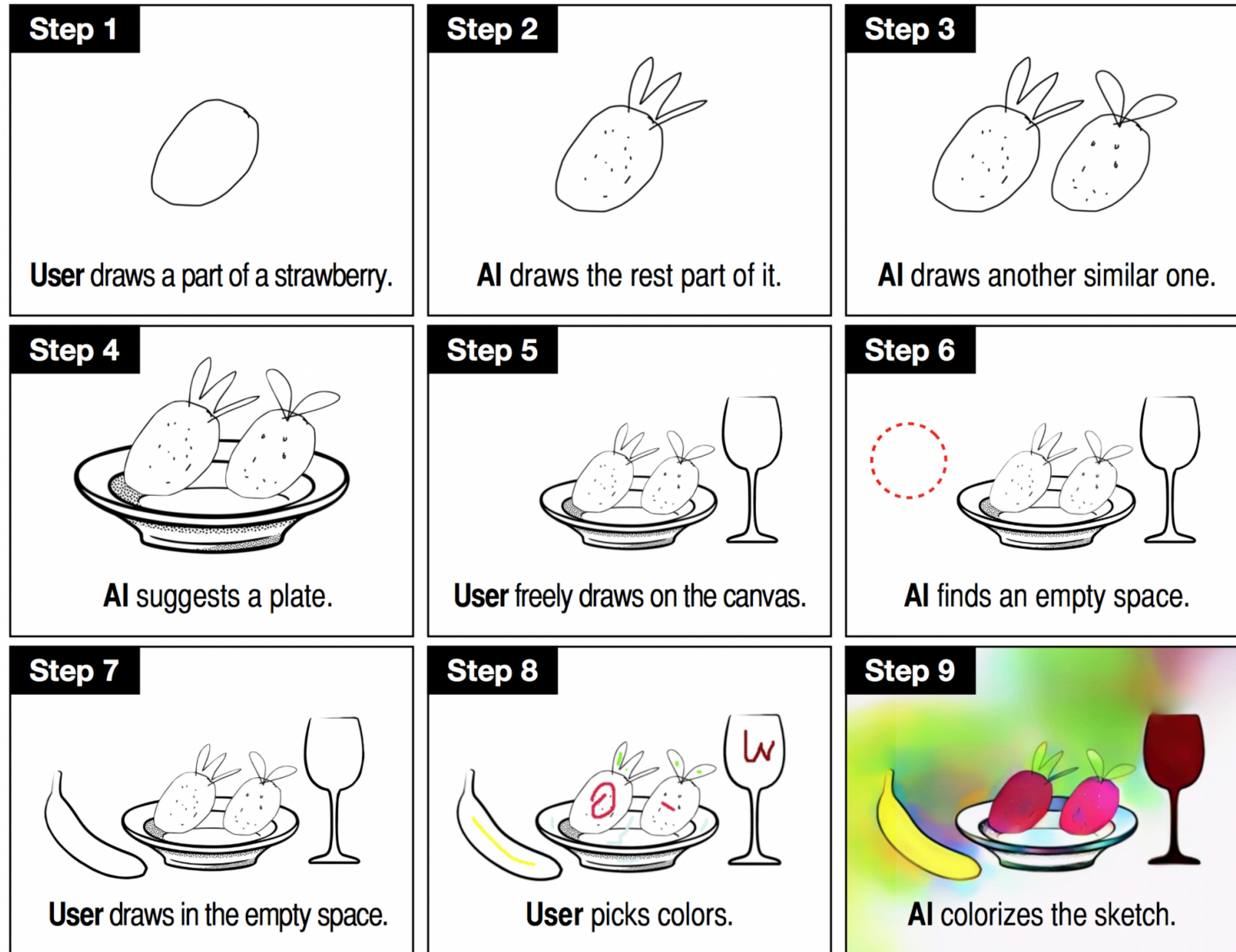
Pedro Uria-Recio, 2019

Human-AI Collaboration

The **cooperative** and **coordinated** interaction between humans (*mostly non-AI experts*) and AI to solve *complex problems* or *achieve certain goals*.

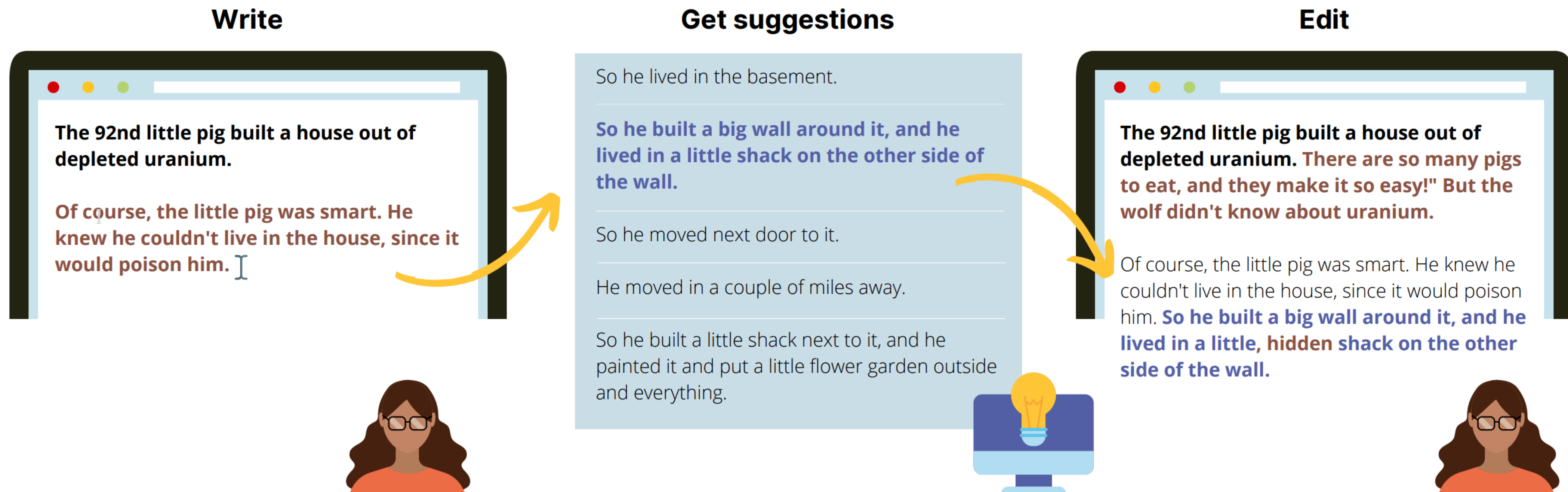
What are some examples?

What are some examples of HAI Collaboration?



Oh, Changhoon, et al. "I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence." CHI 2018

What are some examples of HAI Collaboration?



What are some examples of HAI Collaboration?

```
sentiment.ts  write_sql.go  parse_expenses.py  addresses.rb

1 #!/usr/bin/env ts-node
2
3 import { fetch } from "fetch-h2";
4
5 // Determine whether the sentiment of text is positive
6 // Use a web service
7 async function isPositive(text: string): Promise<boolean> {
8   const response = await fetch(`http://text-processing.com/api/sentiment/`, {
9     method: "POST",
10    body: `text=${text}`,
11    headers: {
12      "Content-Type": "application/x-www-form-urlencoded",
13    },
14  });
15  const json = await response.json();
16  return json.label === "pos";
17 }
```

What are some examples of HAI Collaboration?

I need to throw a dinner party for 6 people who don't eat nuts or seafood. Can you suggest a 3-course menu?

✓ Search the web for: **3 course menu no nuts no seafood**

✓ Search the web for: **course menu examples**

A possible 3-course menu for 6 people who don't like nuts or seafood is:

- **Starter:** Vegetable soup with gluten-free bread. Warm and healthy dish with various veggies and herbs.
- **Main:** Roasted chicken with roasted potatoes and green beans. Classic and satisfying dish with garlic, lemon, and rosemary.
- **Dessert:** Apple crisp with vanilla ice cream. Delicious and nut-free dessert with cinnamon and gluten-free crumble.

I have sweet potato already. Can you make the main course use that?

Sure, you can use sweet potato in your main course. Here is one possible option:

- **Sweet potato and spinach lasagna.** This is a vegetarian and gluten-free dish that layers sweet potato slices, spinach, ricotta cheese, and tomato sauce in a baking dish.

Human-AI Collaboration: key aspects

Complementary performance: Leverage the strengths of both AI and humans, to achieve better outcomes than either could accomplish alone.

Alignment: AI should behave in a ways that humans expect their “teammate” to behave.

Cooperative interaction: The interaction between human and AI should work in a way that’s both intuitive for the human, and compatible to model characteristics.

Reasonable task allocation: Assign tasks that the model can perform.

Appropriate trust: humans should follow AIs when they are correct, but identify and correct their mistakes otherwise.

Human-AI Collaboration: key aspects

Complementary performance: Leverage the strengths of both AI and humans, to achieve better outcomes than either could accomplish alone.

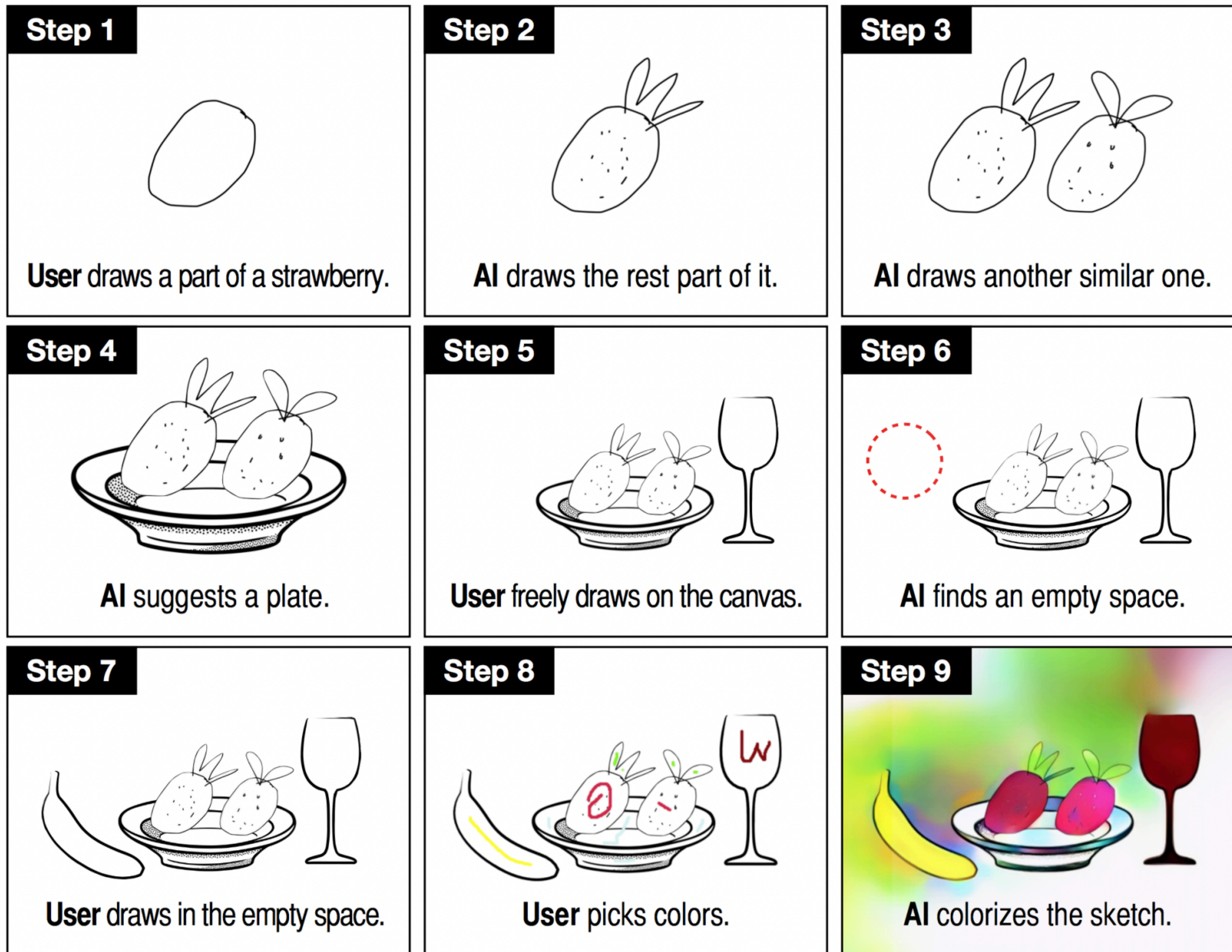
Alignment: AI should behave in a ways that humans expect their “teammate” to behave.

Cooperative interaction: The interaction between human and AI should work in a way that’s both intuitive for the human, and compatible to model characteristics.

Reasonable task allocation: Assign tasks that the model can perform.

Appropriate trust: humans should follow AIs when they are correct, but identify and correct their mistakes otherwise.

What are some examples of HAI Collaboration?



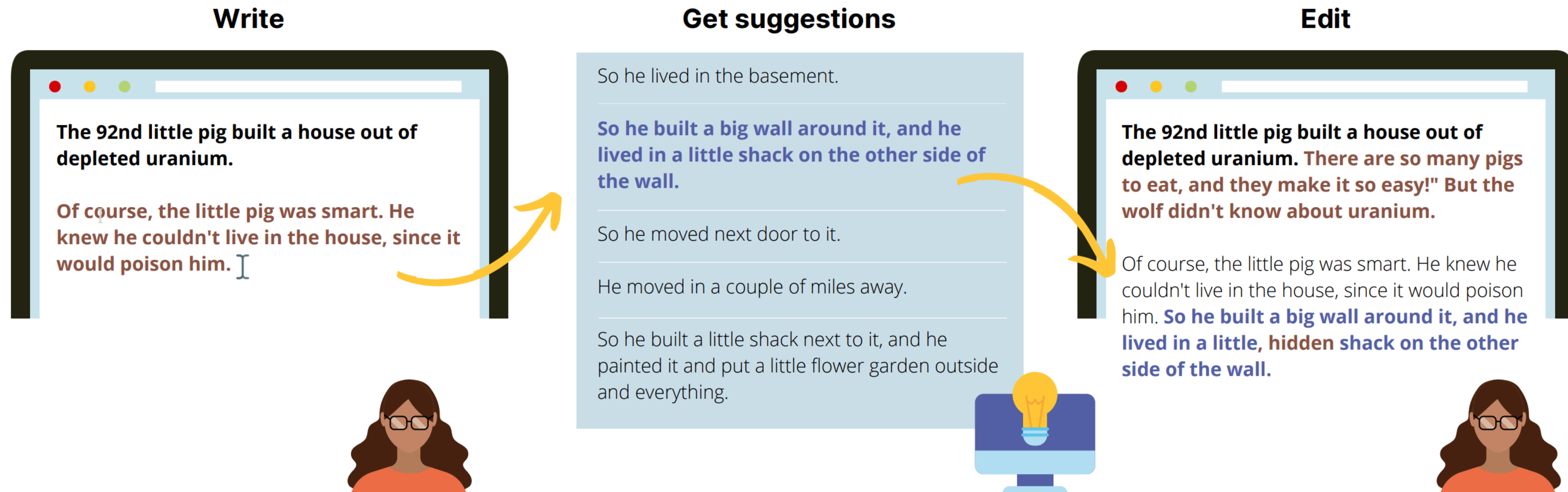
Human (*good at taking visual signals, draw rough shapes, annotate colors*):
lead the creation

Model (*repetition, detect space, detect objects & find similar objects*):

Automate repetitive tasks like draw the dots, duplicate the figure, fill in the color, suggest places to draw more

Oh, Changhoon, et al. "I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence." CHI 2018

What are some examples of HAI Collaboration?



Human (*good at logical reasoning and consistency in long doc, know what they want*)

Lead the writing, edit the model suggestions.

Model (*good at quick generate text many versions of text based on local context*)

Suggest next sentences, help write faster & overcome writer's block

What are some examples of HAI Collaboration?

I need to throw a dinner party for 6 people who don't eat nuts or seafood. Can you suggest a 3-course menu?

✓ Search the web for: **3 course menu no nuts no seafood**

✓ Search the web for: **course menu examples**

A possible 3-course menu for 6 people who don't like nuts or seafood is:

- **Starter:** Vegetable soup with gluten-free bread. Warm and healthy dish with various veggies and herbs.
- **Main:** Roasted chicken with roasted potatoes and green beans. Classic and satisfying dish with garlic, lemon, and rosemary.
- **Dessert:** Apple crisp with vanilla ice cream. Delicious and nut-free dessert with cinnamon and gluten-free crumble.

I have sweet potato already. Can you make the main course use that?

Sure, you can use sweet potato in your main course. Here is one possible option:

- **Sweet potato and spinach lasagna.** This is a vegetarian and gluten-free dish that layers sweet potato slices, spinach, ricotta cheese, and tomato sauce in a baking dish.

Human (*know what they need*) Iteratively refine their search query

Bing search chat (*have access to web*) Provide answer given the search constraint

Human-AI Collaboration: key aspects

Complementary performance: Leverage the strengths of both AI and humans, to achieve better outcomes than either could accomplish alone.

Alignment: AI should behave in a ways that humans expect their “teammate” to behave.

Cooperative interaction: The interaction between human and AI should work in a way that’s both intuitive for the human, and compatible to model characteristics.

Reasonable task allocation: Assign tasks that the model can perform.

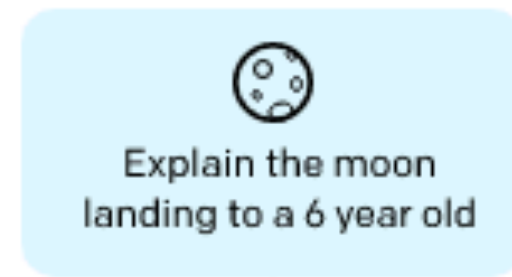
Appropriate trust: humans should follow AIs when they are correct, but identify and correct their mistakes otherwise.

Align models with humans through human feedback

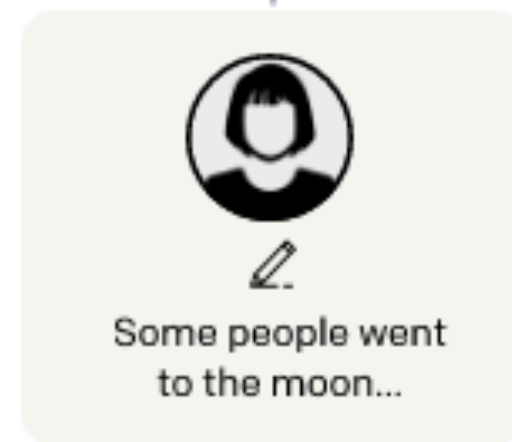
Step 1

Collect demonstration data, and train a supervised policy.

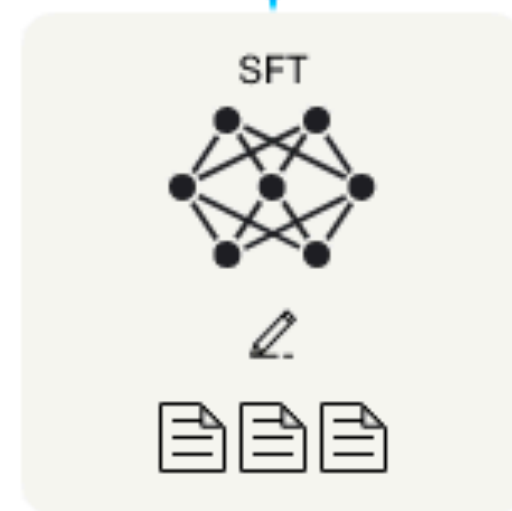
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



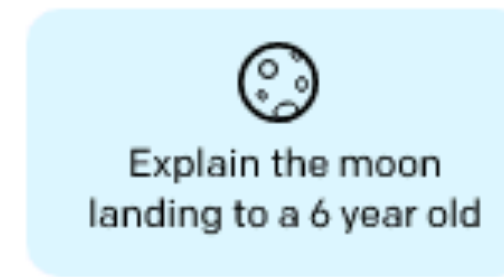
This data is used to fine-tune GPT-3 with supervised learning.



Step 2

Collect comparison data, and train a reward model.

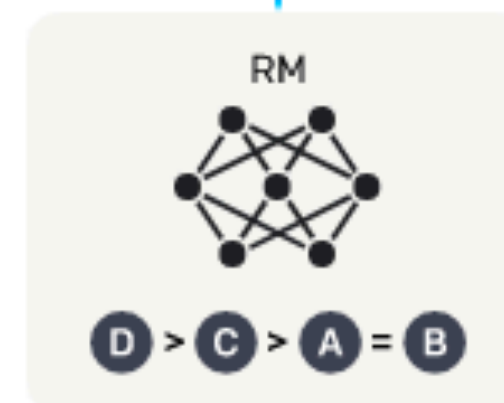
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



Step 3

Optimize a policy against the reward model using reinforcement learning.

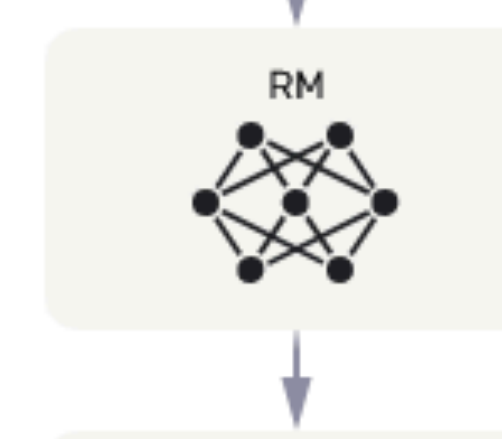
A new prompt is sampled from the dataset.



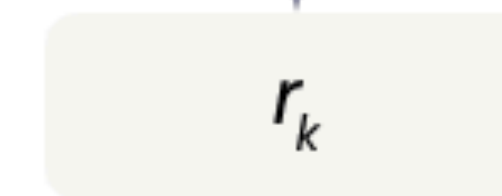
The policy generates an output.



The reward model calculates a reward for the output.



The reward is used to update the policy using PPO.



Work with InstructGPT: Prompt Engineering

Prompts involve instructions and context passed to a LM to achieve a desired task

Prompt engineering is the practice of developing and optimizing prompts to efficiently use language models (LMs) for a variety of applications

A prompt is composed with the following components:

Instructions

```
Classify the text into neutral, negative or positive
```

Input context, & data

```
Text: I think the food was great.
```

Output indicator, & output

```
Sentiment: positive
```

Human-AI Collaboration: key aspects

Complementary performance: Leverage the strengths of both AI and humans, to achieve better outcomes than either could accomplish alone.

Alignment: AI should behave in a ways that humans expect their “teammate” to behave.

Cooperative interaction: The interaction between human and AI should work in a way that’s both intuitive for the human, and compatible to model characteristics.

Reasonable task allocation: Assign tasks that the model can perform.

Appropriate trust: humans should follow AIs when they are correct, but identify and correct their mistakes otherwise.

Work with LLM: Account for model characteristics

Language models are not deterministic.

Language models are lack of reasoning capabilities.

Language models sometimes cannot understand instructions.

Model is not deterministic.

My favorite animal is a dog

How should we deal with such “randomness”?

Depends on the task – remove, express, or exploit!

dog = 8.53%

cat = 5.12%

gir = 4.71%

horse = 3.66%

dolphin = 3.64%

p = 2.92%

pig = 2.75%

lion = 2.26%

tiger = 2.21%

pengu = 2.17%

Total: -2.46 logprob on 1 tokens
(37.96% probability covered in top 10 logits)

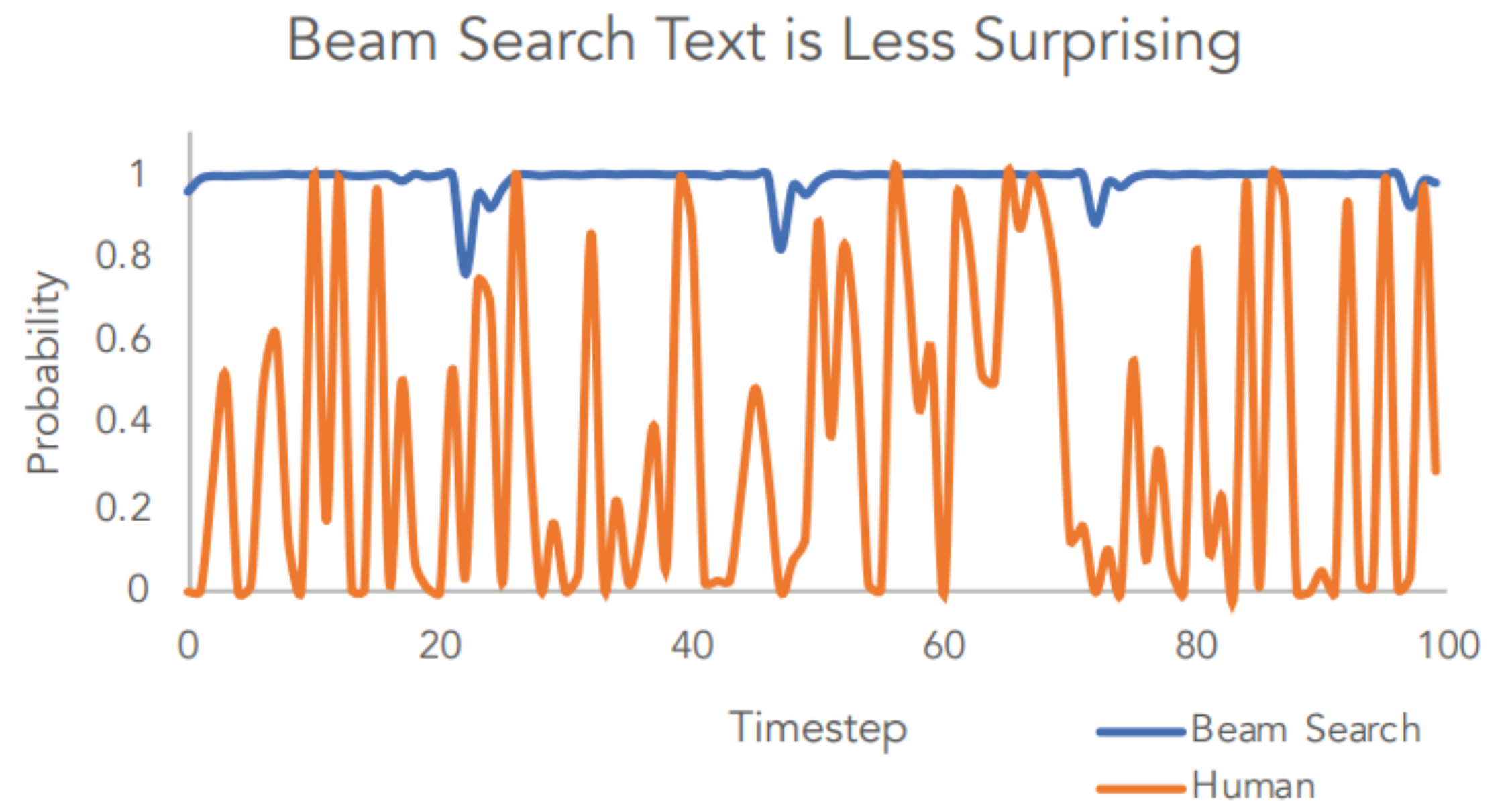
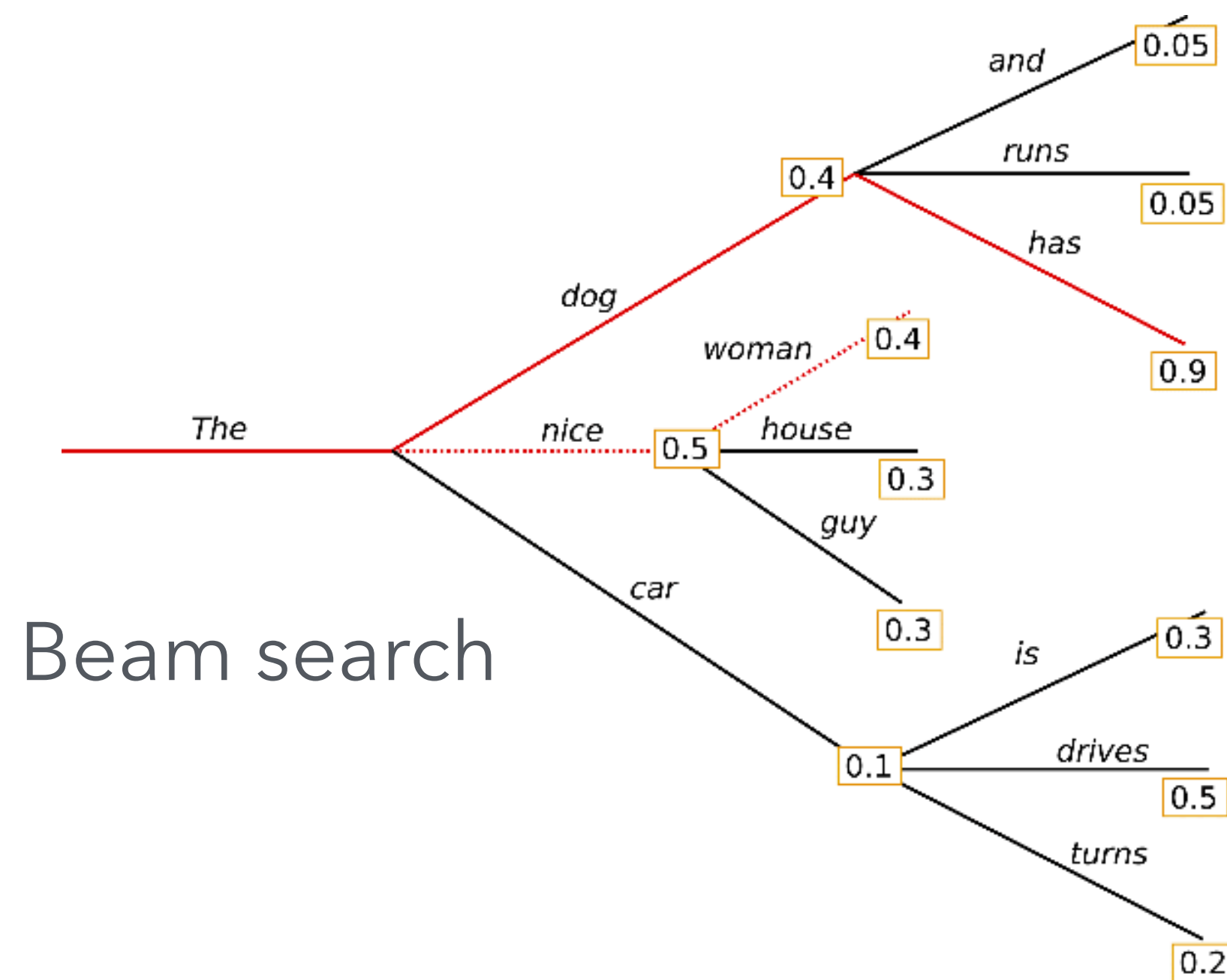
Non-deterministic LMs → Remove uncertainty

When tasks need certainty,

(e.g. write formal documents, need to maximize grammatical correctness, do classification)

remove uncertainty through:

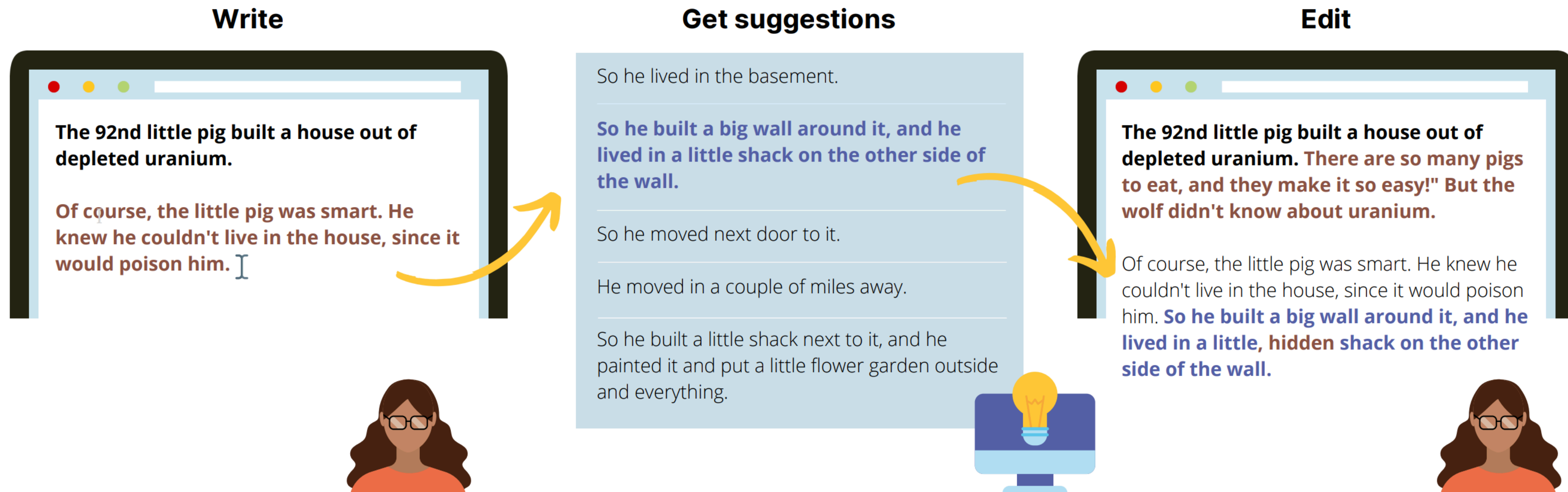
parameters (temperature=0), **less surprising sampling** (e.g. beam search or look ahead)



<https://huggingface.co/blog/how-to-generate>

Non-deterministic LMs Exploit uncertainty

When tasks need creativity,
(e.g. creative writing, ideation, etc.)
exploit uncertainty through:
parameters (temperature=1), re-running the generation multiple times



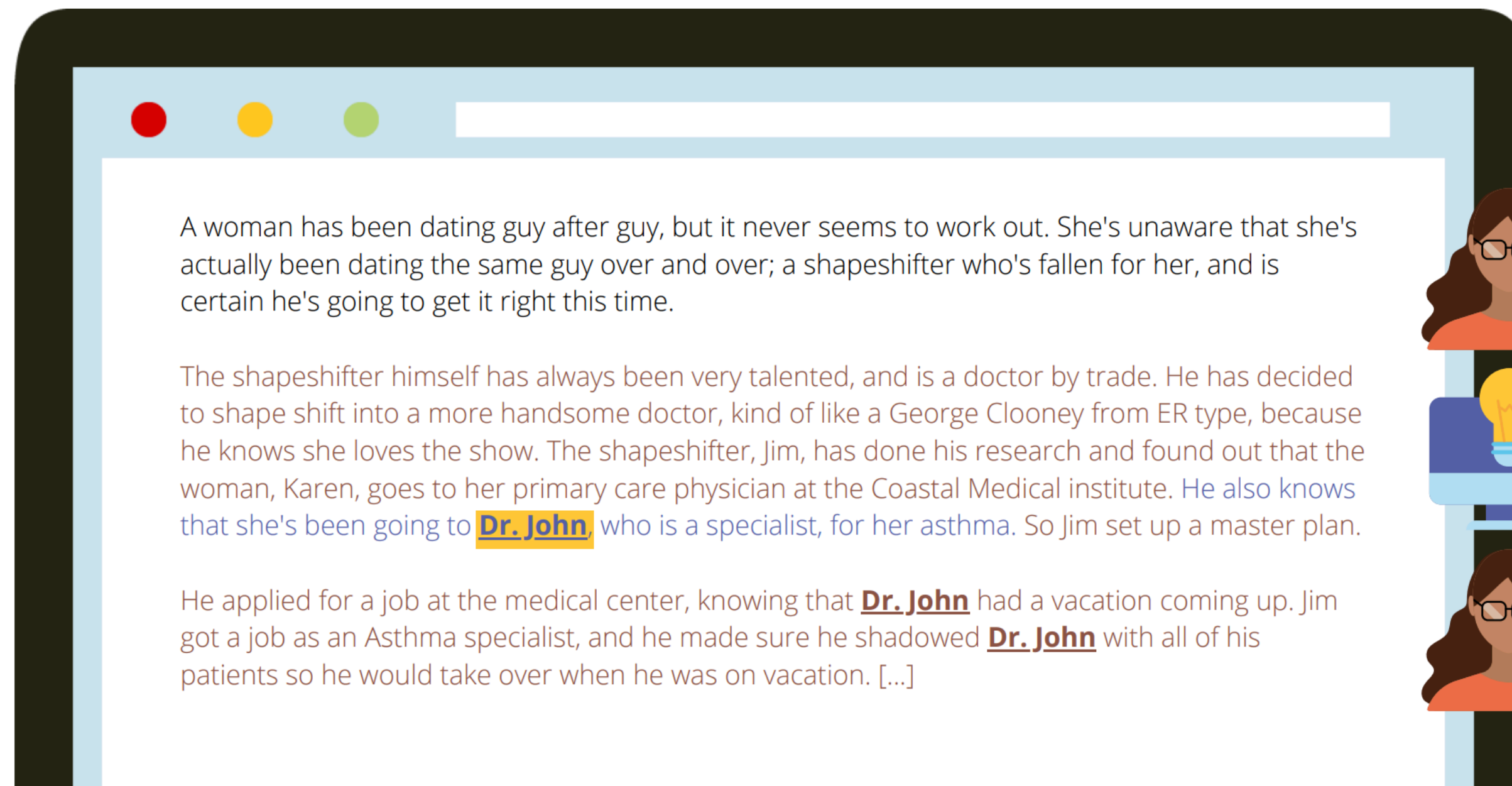
Non-deterministic LMs Exploit uncertainty

When tasks need creativity,

(e.g. creative writing, ideation, etc.)

exploit uncertainty through:

parameters (temperature=1), re-running the generation multiple times



“Model hallucination is not a bug but a feature”: Can be **inspirations** to humans.

Lack-of-reasoning → Chain-of-Thought

Plain prompting

Model cannot do reasoning based on word distribution

I went to the market and bought 10 apples. I gave 2 apples to the neighbor and 2 to the repairman. I then went and bought 5 more apples and ate 1. How many apples did I remain with?

11 apples

CoT prompting

With "step-by-step", can make the final output depend on intermediate ones

I went to the market and bought 10 apples. I gave 2 apples to the neighbor and 2 to the repairman. I then went and bought 5 more apples and ate 1. How many apples did I remain with?

Let's think step by step.

First, you started with 10 apples. You gave away 2 apples to the neighbor and 2 to the repairman, so you had 6 apples left. Then you bought 5 more apples, so now you had 11 apples. Finally, you ate 1 apple, so you would remain with 10 apples.

Prompting can be further improved by *instructing the model to reason* about the task.

Cannot understand instruction → Few-shot

Zero-shot

Natural language descriptions only

1 Find the nationality of people: — Task instruction
2 Marie Curie => — Task

Few-shot

Description + a few example (3-100)
[5-10 is most common]

1 Find the nationality of people: — Task instruction
2 Albert Einstein => German — Examples
3 Alan Turing => English — Examples
4 Mahatma Gandhi => Indian — Examples
5 Marie Curie => — Task

ChatGPT vs. **InstructGPT**: Different Interface

The screenshot shows the OpenAI Playground interface. The browser address bar displays `beta.openai.com/playground/p/L5FVM9jP9sGLVcH2tGweZI5C`. The page header includes the OpenAI logo, a 'Beta' badge, and navigation links for 'Playground', 'Documentation', and 'Examples'. A user profile 'Personal' is visible in the top right.

The main content area is titled 'Playground' and contains a text prompt: "For text generation text using AI, OpenAI's third-generation transformer, GPT-3, is currently state-of-the-art. It can generate high-quality text given several examples, but I wanted to see if I could prompt the system to adopt the writing style from one writer and transfer it to another writer's work. For this experiment, which I call VoxMorphia, I chose to transform songs that are in the public domain. They include classic songs like 'Yes Sir, That's My Baby', with lyrics by Gus Kahn from 1925. Using GPT-3 and some additional open-source projects, I transferred the lyrical style of Bob Marley, Thom Yorke from Radiohead, Megan Thee Stallion, and others to update the words of some classic tunes. The heavy lifting for this project is done by GPT-3 to generate the lyrics and Google's Universal Sentence Encoder to help with pre and post-processing the text. TL;DR This project is a proof-of-concept that AI can be used to generate new text based on the style of another writer. This shows that AI can be used to generate new content and that AI-generated text can be used in creative ways|".

Below the prompt is a 'Submit' button and navigation icons. To the right, the 'Engine' settings are visible: 'davinci' is selected in the dropdown, 'Response Length' is set to 256, 'Temperature' is 0.5, and 'Top P' is 1.

```
Example request      text-davinci-003 v python v Copy

1 import os
2 import openai
3 openai.api_key = os.getenv("OPENAI_API_KEY")
4 openai.Completion.create(
5     model="text-davinci-003",
6     prompt="Say this is a test",
7     max_tokens=7,
8     temperature=0
9 )
```

ChatGPT vs. InstructGPT: Different Interface

SYSTEM
You are an AI research assistant. You use a tone that is technical and scientific.

USER Hello, who are you?

ASSISTANT Greeting! I am an AI research assistant. How can I help you today?

USER Can you tell me about the creation of black holes?

ASSISTANT Of course! Black holes are created when a massive star runs out of fuel and collapses under its own gravity. This collapse causes the star to become incredibly dense, with a gravitational field so strong that not even light can escape it. This point of no return is known as the event horizon. Once a star crosses the event horizon, it is considered a black hole.

There are also intermediate black holes, which are believed to form from the merging of small black holes. Supermassive black holes, which can be billions of times the mass of the sun, are thought to form through a combination of different processes, including the accretion of gas and the merging of smaller black holes.

Mode
Chat Beta

Model
gpt-3.5-turbo

Temperature 0.7

Maximum length 256

Top P 1

Frequency penalty 0

Presence penalty 0

Submit

```
import openai

openai.ChatCompletion.create(
    model="gpt-3.5-turbo",
    messages=[
        {"role": "system", "content": "You are an AI research assistant. You use a tone th"},
        {"role": "user", "content": "Hello, who are you?"},
        {"role": "assistant", "content": "Greeting! I am an AI research assistant. How car"},
        {"role": "user", "content": "Can you tell me about the creation of black holes?"}
    ]
)
```

Interface changes how you use the AI!

Give me a workout I can do at home

✓ Generating answers for you...

Sure, I can help you with that. There are many exercises you can do at home without any equipment, such as **jumping jacks**, **burpees**, **squats**, **push-ups**, **lunges**, and more. Here is a sample workout routine you can try at home, based on the Nerd Fitness website:

- Warm up for 5 to 10 minutes with some light cardio and stretching.
- Do 20 bodyweight squats.
- Do 10 lunges on each leg.

Ask me anything



Interface changes how you use the AI!

Give me a workout I can do at home

✓ Generating answers for you...

Sure, I can help you with that. There are many exercises you can do at home without any equipment, such as **jumping jacks, burpees, squats, push-ups, lunges**, and more. Here is a sample workout routine you can try at home, based on the Nerd Fitness website:

- Warm up for 5 to 10 minutes with some light cardio and stretching.
- Do 20 bodyweight squats.
- Do 10 lunges on each leg.

Can you focus on arms

✓ Generating answers for you...



Ask me anything



e.g., Compared to traditional search, Bing w/ ChatGPT is more **conversational** and **interactive**...






You use more **complete sentences**.

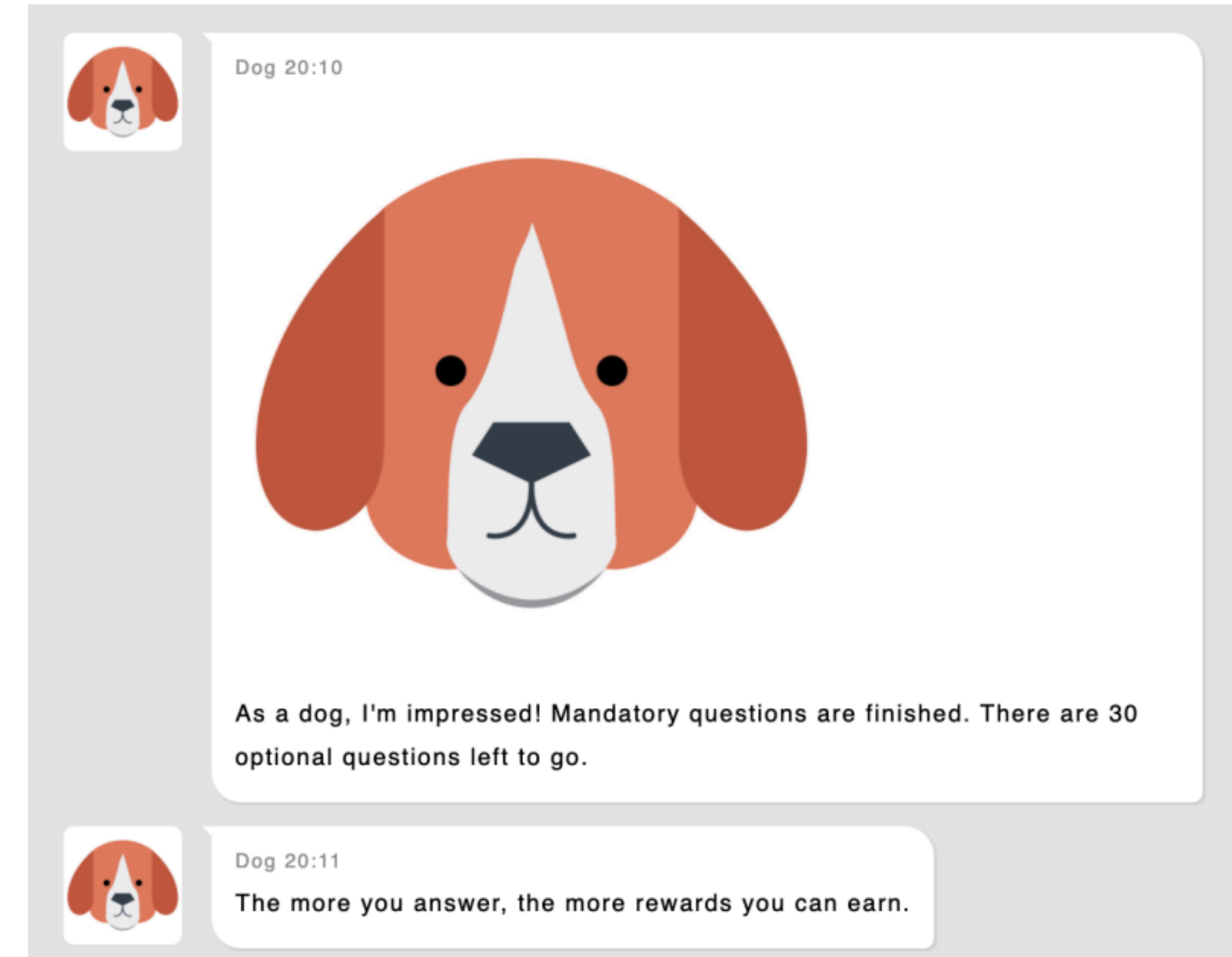
You get a sense of talking to someone and therefore become more **polite** ("can you...")

You can **follow up** more naturally, in a context-aware way.

You use more "**engaging**" prompts, e.g. ask questions, provide commands, do reflections, etc.

Higher level: Human perception on AI change things

| Opening greetings | |
|---|---|
|  God | Greetings, I am the lord thy god. You shall help researchers at a university by participating in this research. So I ask you, does this sound good to you? |
|  Human | Hello, pleasure to meet you! I'm a trained professional who will guide you here. Researchers in a university asked me to facilitate this research. Does this sound good to you? |
|  Animal | Woof woof, nice to meet you! I'm a dog that will help you conduct this research (...) |
|  Plant | Hello, nice to meet you! I'm an avocado that will help you conduct this research (...) |
|  Inorganic Object | Hello, nice to meet you! I'm a book that will help you conduct this research (...) |
| Control | Hello! Can you help researchers in a university by conducting this research? |



| | User Engagement Scale | Perceived Usability (UES) | Aesthetic Appeal (UES) | Interest-Enjoyment (IMI) | Cognitive Load |
|----------------|-----------------------|---------------------------|------------------------|-------------------------------|--------------------------|
| God | - | ↓ than Control | - | - | ↑ than Object, Control |
| Human | - | - | - | ↑ than Object | ↑ than Object |
| Animal | ↑ than Object | - | ↑ than Object | ↑ than Object | - |
| Plant | - | - | - | - | ↑ than Object |
| Object | ↓ than Animal | - | ↓ than Animal | ↓ than Human, Animal, Control | ↓ than God, Human, Plant |
| Control | - | ↑ than God | - | ↑ than Object | ↓ than God |

Higher level: Human perception on AI change things



In other words, when designing interactions, we should consider what the model can do, and also manipulate how we want humans to think of AI models depending on our goals.

| | User Engagement Scale | Perceived Usability (UES) | Aesthetic Appeal (UES) | Interest-Enjoyment (IMI) | Cognitive Load |
|----------------|-----------------------|---------------------------|------------------------|-------------------------------|--------------------------|
| God | - | ↓ than Control | - | - | ↑ than Object, Control |
| Human | - | - | - | ↑ than Object | ↑ than Object |
| Animal | ↑ than Object | - | ↑ than Object | ↑ than Object | - |
| Plant | - | - | - | - | ↑ than Object |
| Object | ↓ than Animal | - | ↓ than Animal | ↓ than Human, Animal, Control | ↓ than God, Human, Plant |
| Control | - | ↑ than God | - | ↑ than Object | ↓ than God |

Human-AI Collaboration: key aspects

Complementary performance: Leverage the strengths of both AI and humans, to achieve better outcomes than either could accomplish alone.

Alignment: AI should behave in a ways that humans expect their “teammate” to behave.

Cooperative interaction: The interaction between human and AI should work in a way that’s both intuitive for the human, and compatible to model characteristics.

Reasonable task allocation: Assign tasks that the model can perform.

Appropriate trust: humans should follow AIs when they are correct, but identify and correct their mistakes otherwise.

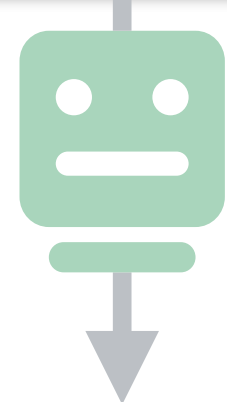
A case study: LLMs as multiple teammates.

Single models *struggle* with *complex* tasks. Example: Peer review rewriting

Rewriting

Original feedback

Sherry could improve her presentation...



More friendly feedback

Sherry, you have too many words on your...

Given the following feedback, rewrite it into a friendly paragraph with concrete suggestions for each of Alex's presentation problems.

Original feedback:

Sherry could improve her presentation skills. She has too much text on her slides. Her presentation meanders from topic to topic without a clear structure. She also does not engage with her audience when she presents.

More friendly feedback:

Sherry, you have too many words on your slides. You should use images and bullet points to help get your message across. You should have a clear structure for the presentation. You should also engage with your audience.

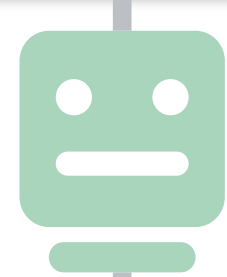
The input-output mapping is **convoluted**.

Example: Peer review rewriting

Rewriting

Original feedback

Sherry could improve her presentation...



Confusing mapping!

More friendly feedback

Sherry, you have too many words on your...

Given the following feedback, rewrite it into a friendly paragraph with concrete suggestions for each of Alex's presentation problems.

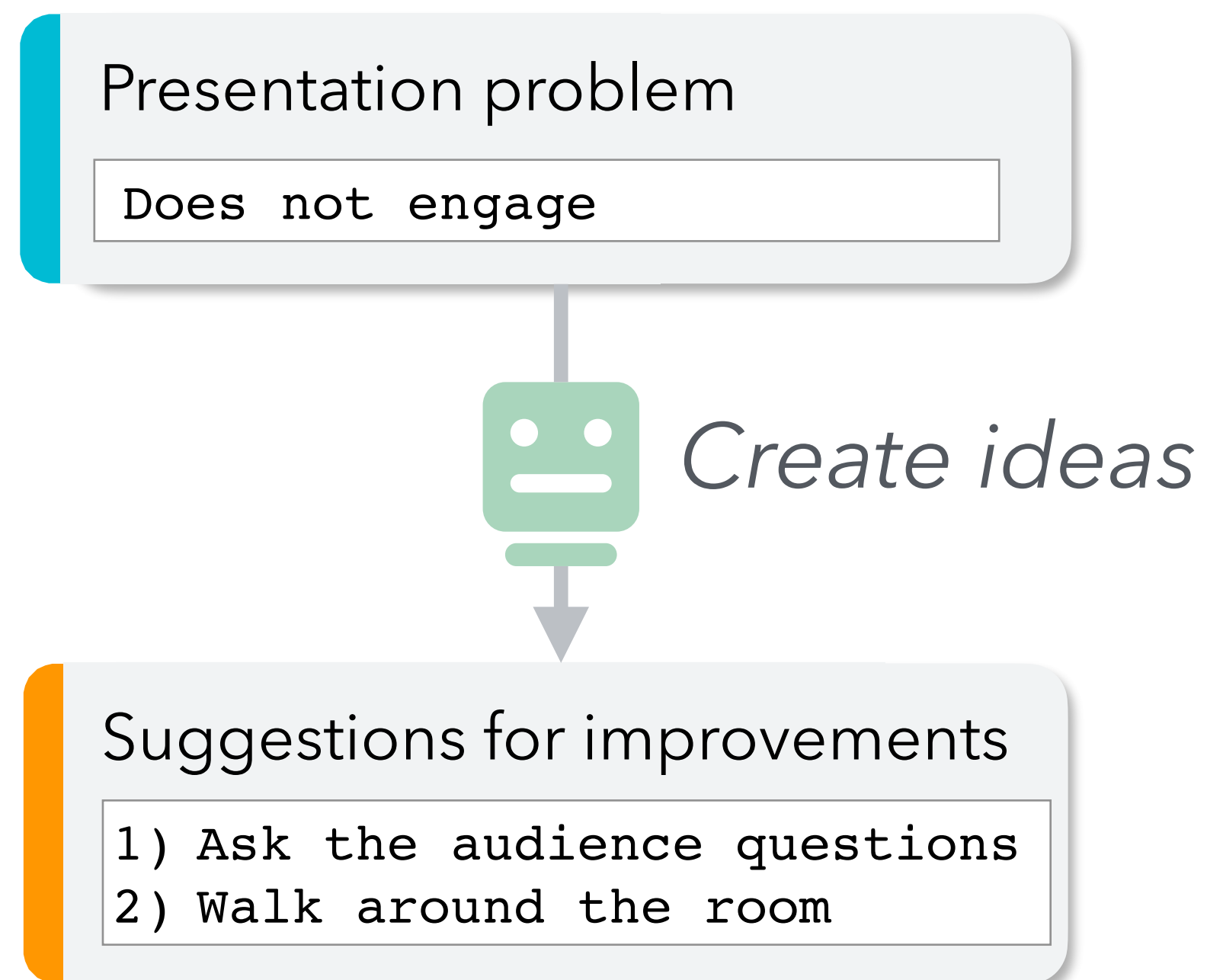
Original feedback:

Sherry could improve her presentation skills. She has too much text on her slides. Her presentation meanders from topic to topic without a clear structure. She also does not engage with her audience when she presents.

More friendly feedback:

Sherry, you have too many words on your slides. You should use images and bullet points to help get your message across. You should have a clear structure for the presentation. You should also engage with your audience.

Small tasks are more **interpretable** and **controllable**.



Given the Presentation problem, the following is a list of improvement suggestions.

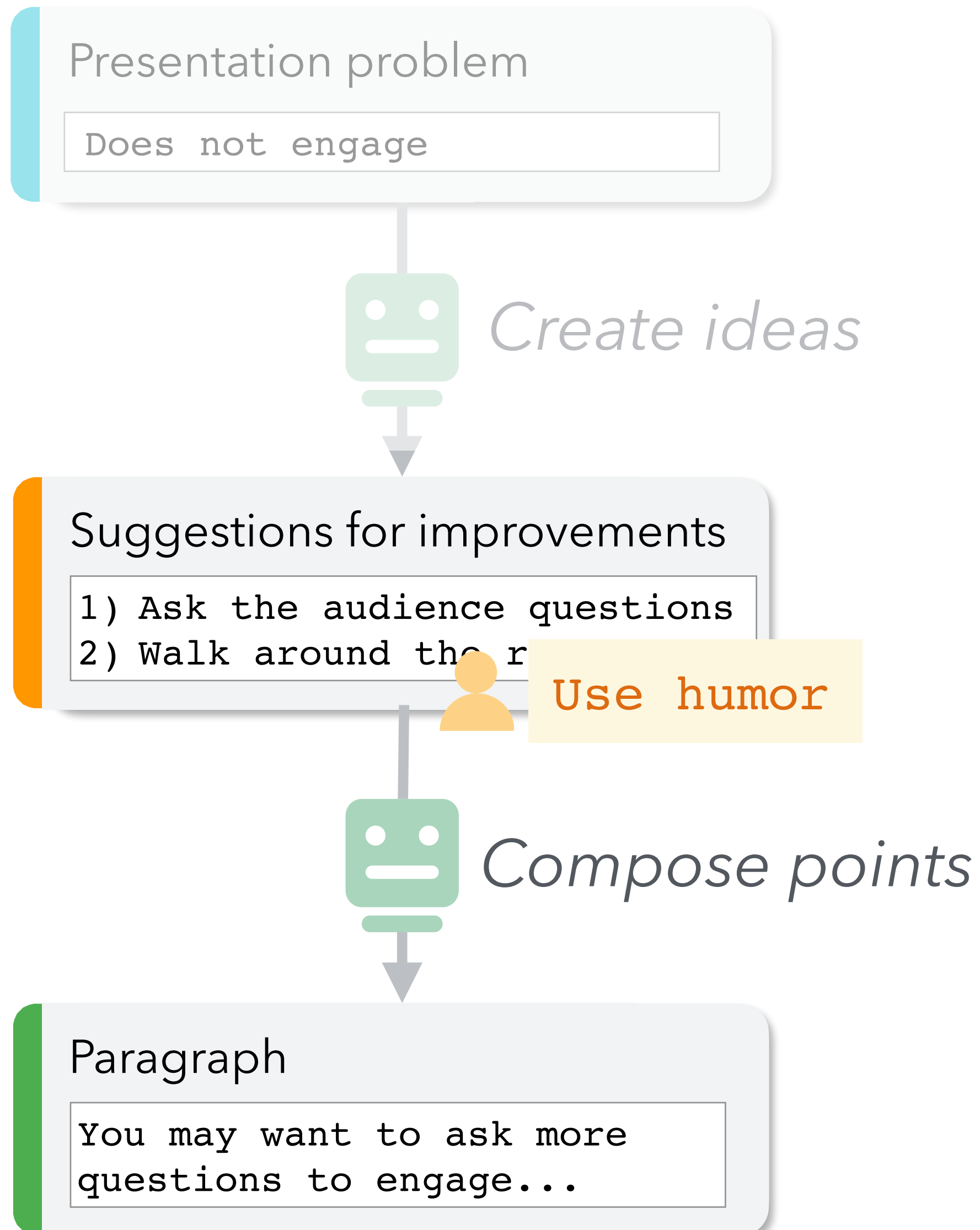
Problem: Does not engage

Suggestions for improvements:

1) Ask the audience questions

 Use humor

The fix can be **propagated** to related sub-tasks!



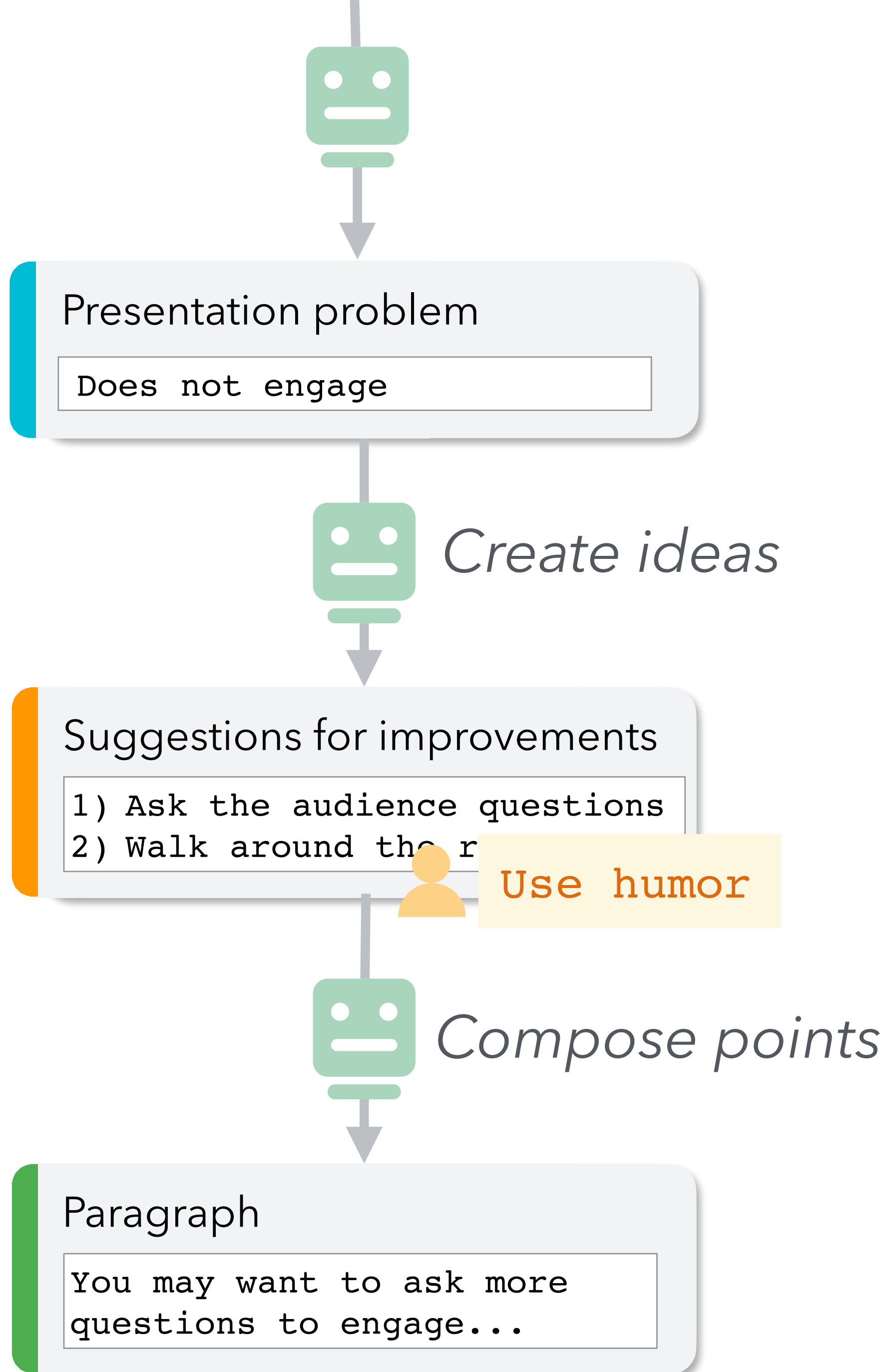
Write one friendly paragraph to cover all the suggestions.

Suggestions:

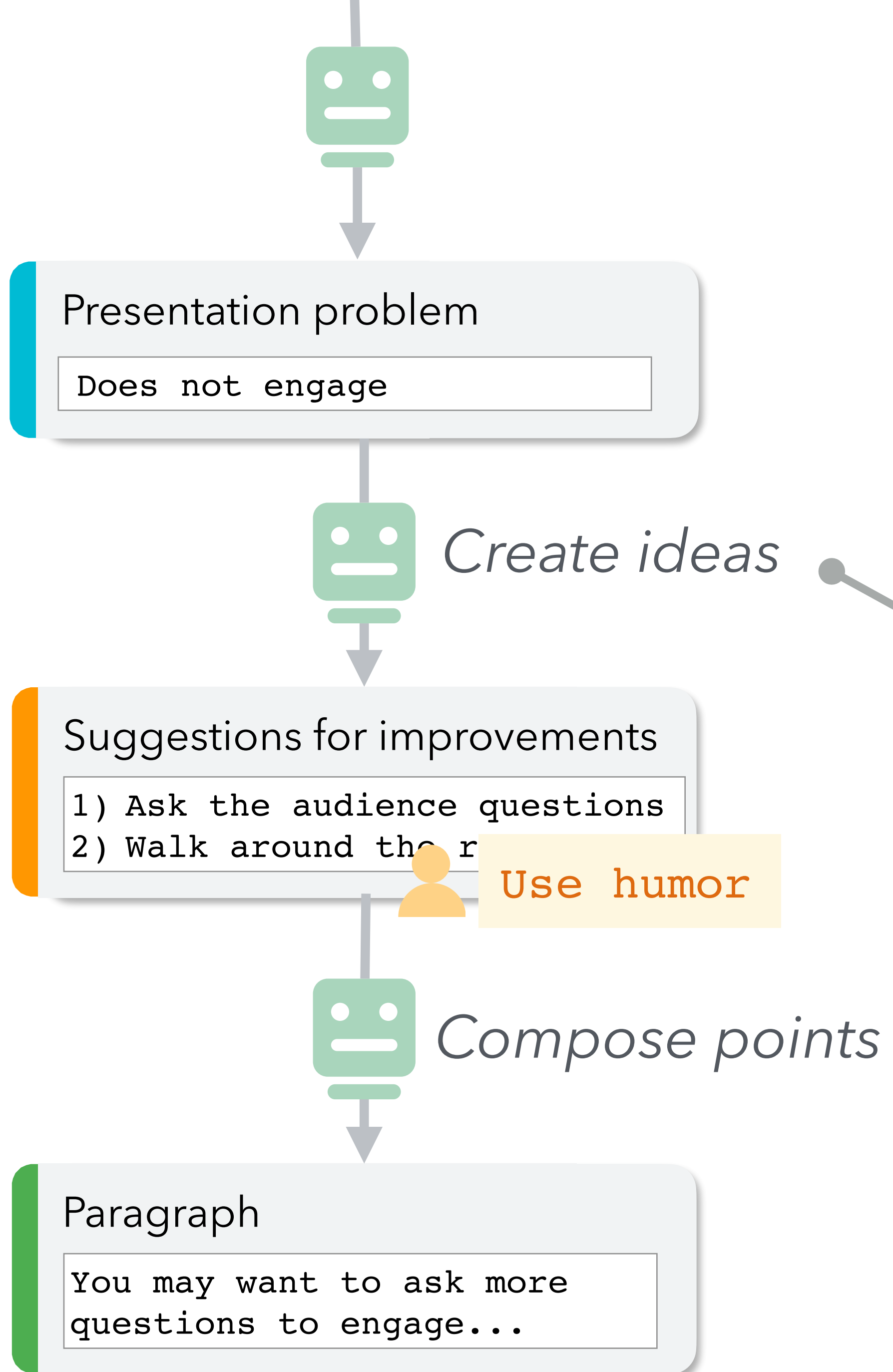
- 1) Ask the audience questions
- 2) Use humor

Paragraph

You may want to ask more questions to engage the audience. Humor always helps!



Chain =
Independent LLM runs per sub-task
+ inter-step transformation



Chain =

Independent LLM runs per sub-task

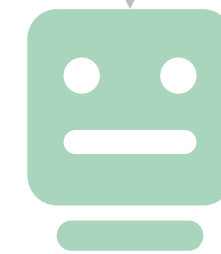
- 1 Identify all presentation problems
- 2 Ideate suggestions per problem
- 3 Compose them into a paragraph

+ inter-step transformation

A three-step LLM Chain for peer review rewriting

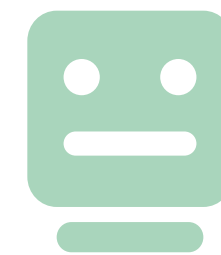
Original feedback

Sherry could improve his presentation skills. He has too much text on his slides. His presentation meanders from topic to topic without a clear structure. He also does not engage with his audience when he presents.



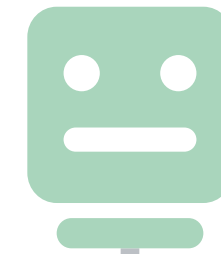
1

Identify all presentation problems



2

Ideate suggestions per problem



3

Compose them into a paragraph

Friendly paragraph

A three-step LLM Chain for peer review rewriting

Original feedback

Sherry could improve his presentation skills. He has too much text on his slides. His presentation meanders from topic to topic without a clear structure. He also does not engage with his audience when he presents.

Sherry's problems

Too much text on slides

No clear structure

Does not engage with audience



1

Identify all presentation problems

Friendly paragraph

A three-step LLM Chain for peer review rewriting

Original feedback

Sherry could improve his presentation skills. He has too much text on his slides. His presentation meanders from topic to topic without a clear structure. He also does not engage with his audience when he presents.

Sherry's problems

Too much text on slides

No clear structure

Does not engage with audience

Ideate suggestions per problem **2**

Friendly paragraph

A three-step LLM Chain for peer review rewriting

Original feedback

Sherry could improve his presentation skills. He has too much text on his slides. His presentation meanders from topic to topic without a clear structure. He also does not engage with his audience when he presents.

Sherry's problems

Too much text on slides

No clear structure

Does not engage with audience

Ideate suggestions per problem **2**

Suggestions for improvement

More images on the slides

Sectionalize the talk

Ask the audience questions

Use humor

Friendly paragraph

A three-step LLM Chain for peer review rewriting

Original feedback

Sherry could improve his presentation skills. He has too much text on his slides. His presentation meanders from topic to topic without a clear structure. He also does not engage with his audience when he presents.

Sherry's problems

Too much text on slides

No clear structure

Does not engage with audience

Suggestions for improvement

More images on the slides

Sectionalize the talk

Ask the audience questions

Use humor

Compose them into a paragraph

3

Original feedback

A three-step LLM Chain for peer review rewriting

Original feedback

Sherry could improve his presentation skills. He has too much text on his slides. His presentation meanders from topic to topic without a clear structure. He also does not engage with his audience when he presents.

Sherry's problems

Too much text on slides

No clear structure

Does not engage with audience

Suggestions for improvement

More images on the slides

Sectionalize the talk

Ask the audience questions

Use humor

Friendly paragraph

Sherry, your presentation was interesting! However, I noticed that you have a lot of information on your slides. It might be helpful to vary pictures with text so that it is easier to follow. Also, you might consider the flow of your theme. If it were me, I would have divided it into three sections and then used your conclusion. You may also want to add some humor, and ask more questions to engage the audience.

Chaining in Open sourcing LangChain

```
from langchain.prompts import PromptTemplate
from langchain.llms import OpenAI

llm = OpenAI(temperature=0.9)
prompt = PromptTemplate(
    input_variables=["product"],
    template="What is a good name for a company that makes {product}?",
)
```

```
second_prompt = PromptTemplate(
    input_variables=["company_name"],
    template="Write a catchphrase for the following company: {company_name}",
)
chain_two = LLMChain(llm=llm, prompt=second_prompt)
```

```
from langchain.chains import SimpleSequentialChain
overall_chain = SimpleSequentialChain(chains=[chain, chain_two], verbose=True)

# Run the chain specifying only the input variable for the first chain.
catchphrase = overall_chain.run("colorful socks")
print(catchphrase)
```

[Read more on
LangChain
documentation](#)

Chaining in OpenSourcing LangChain

```
llm = OpenAI(temperature=0.9)
prompt = PromptTemplate(
    input_variables=["product"],
    template="What is a good name for a company that makes {product}?",
)
```

```
second_prompt = PromptTemplate(
    input_variables=["company_name"],
    template="Write a catchphrase for the following company: {company_name}",
)
chain_two = LLMChain(llm=llm, prompt=second_prompt)
```

```
> Entering new SimpleSequentialChain chain...
```

Cheerful Toes.

"Spread smiles from your toes!"

```
> Finished SimpleSequentialChain chain.
```

```
"Spread smiles from your toes!"
```

[Read more on
LangChain
documentation](#)

Chaining reminds
us of workflows in
crowdsourcing...

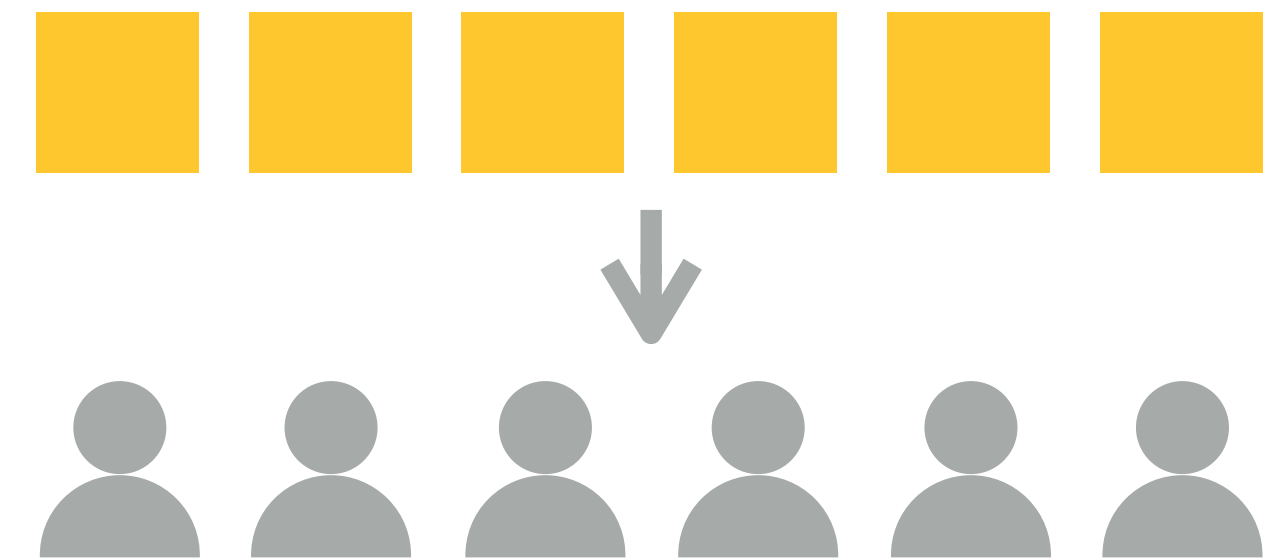
Crowdsourcing: Small Tasks, Many People

Combine many small tasks completed by independent workers.

e.g., text shortening

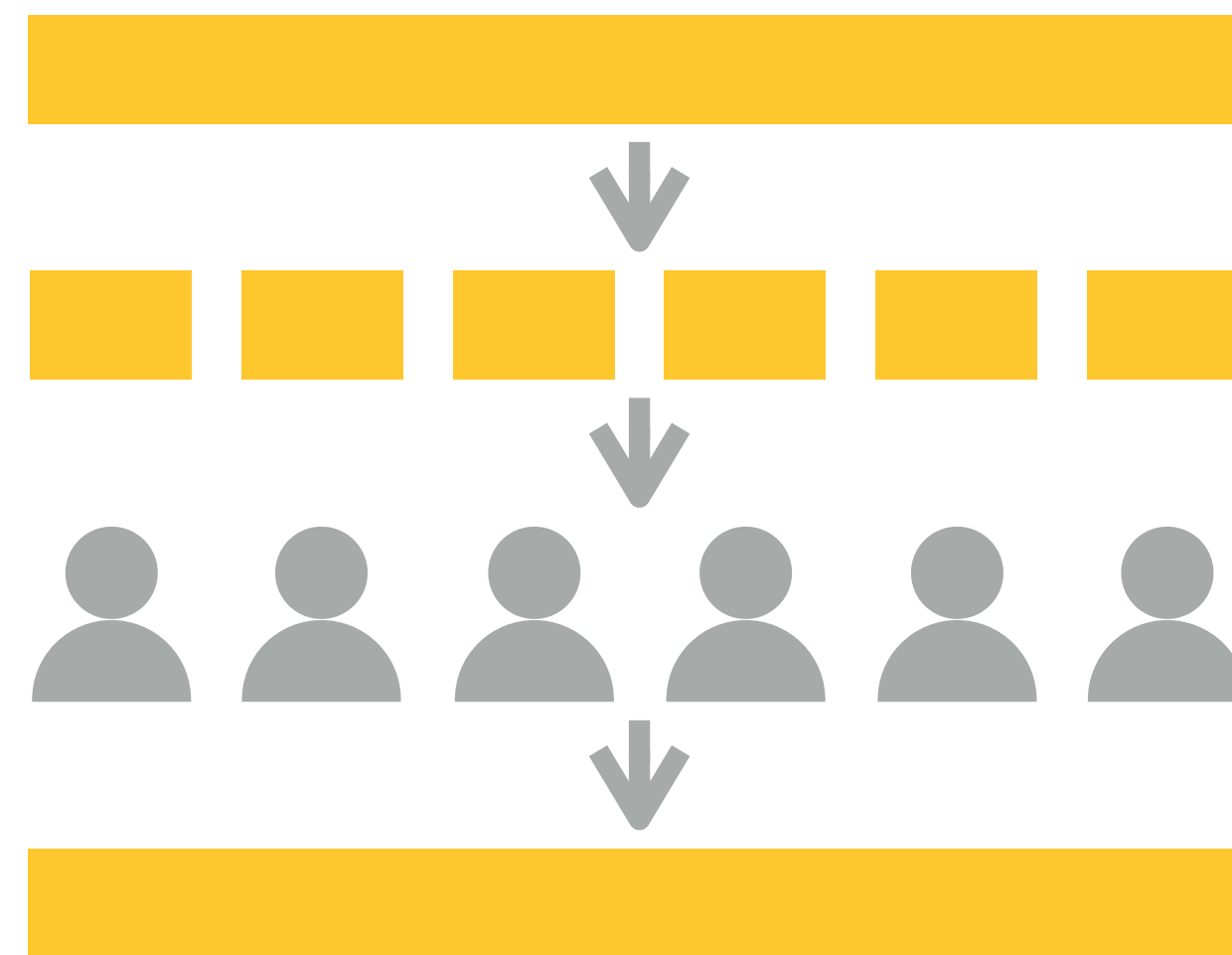
e.g., image labeling

e.g., translation

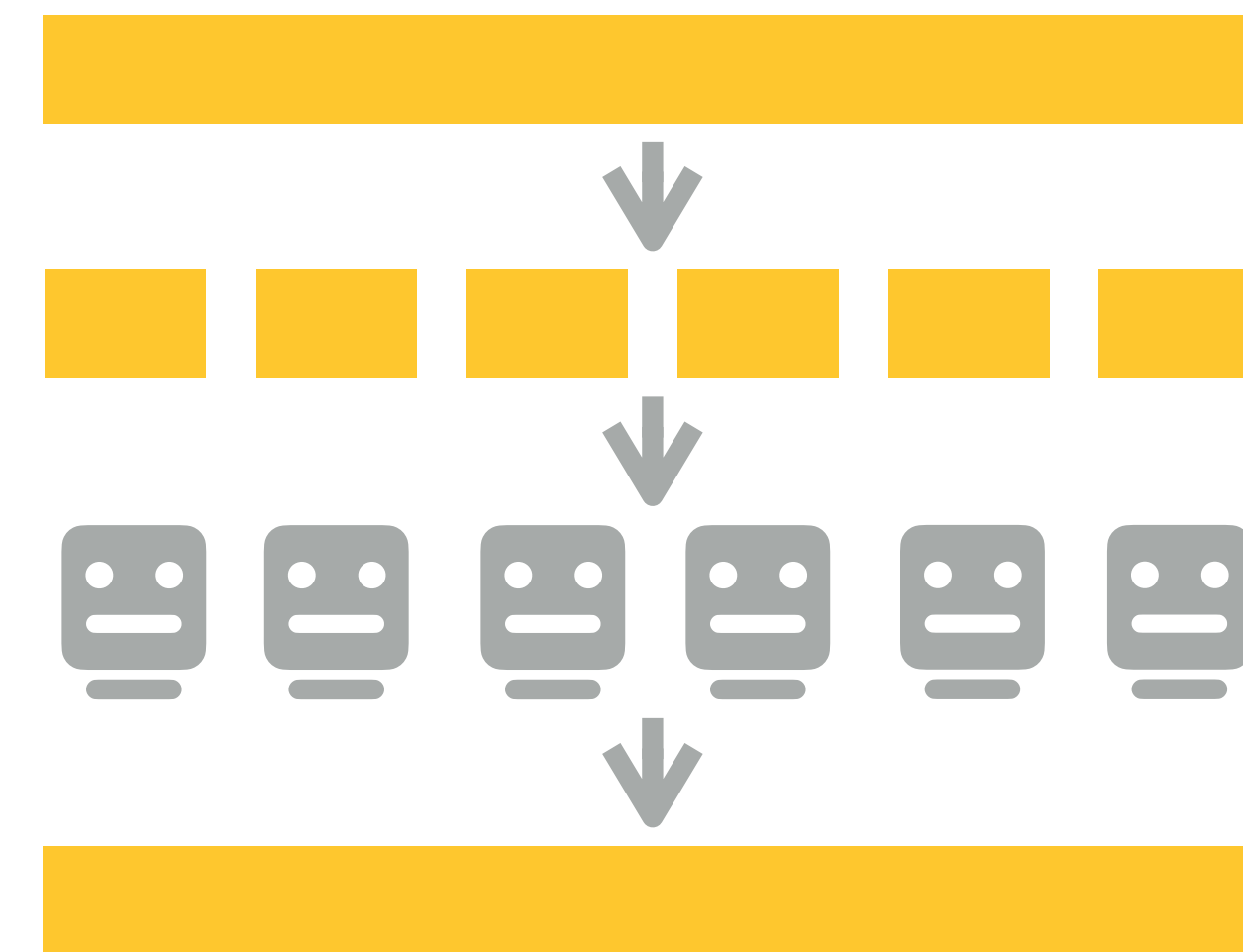


Crowdsourcing workflow vs. Chaining in LLMs

Crowdsourcing workflow



LLM chain



Similarities...

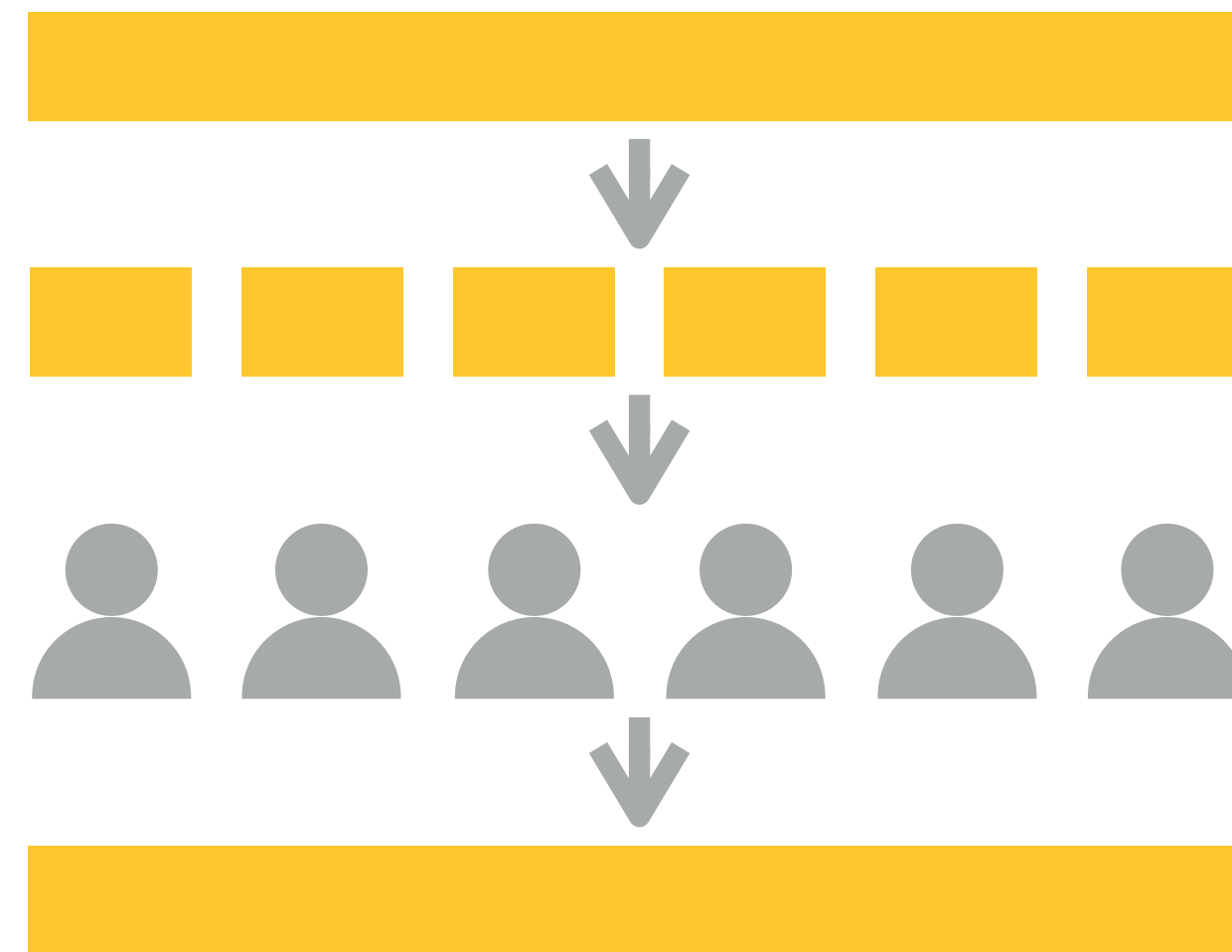
Idea: breakdown complex tasks into pieces that can be done independently, then combined.

Gains: scale to tasks that are otherwise hard, more structured interactions, more resilient to interruptions (of distractor tasks).

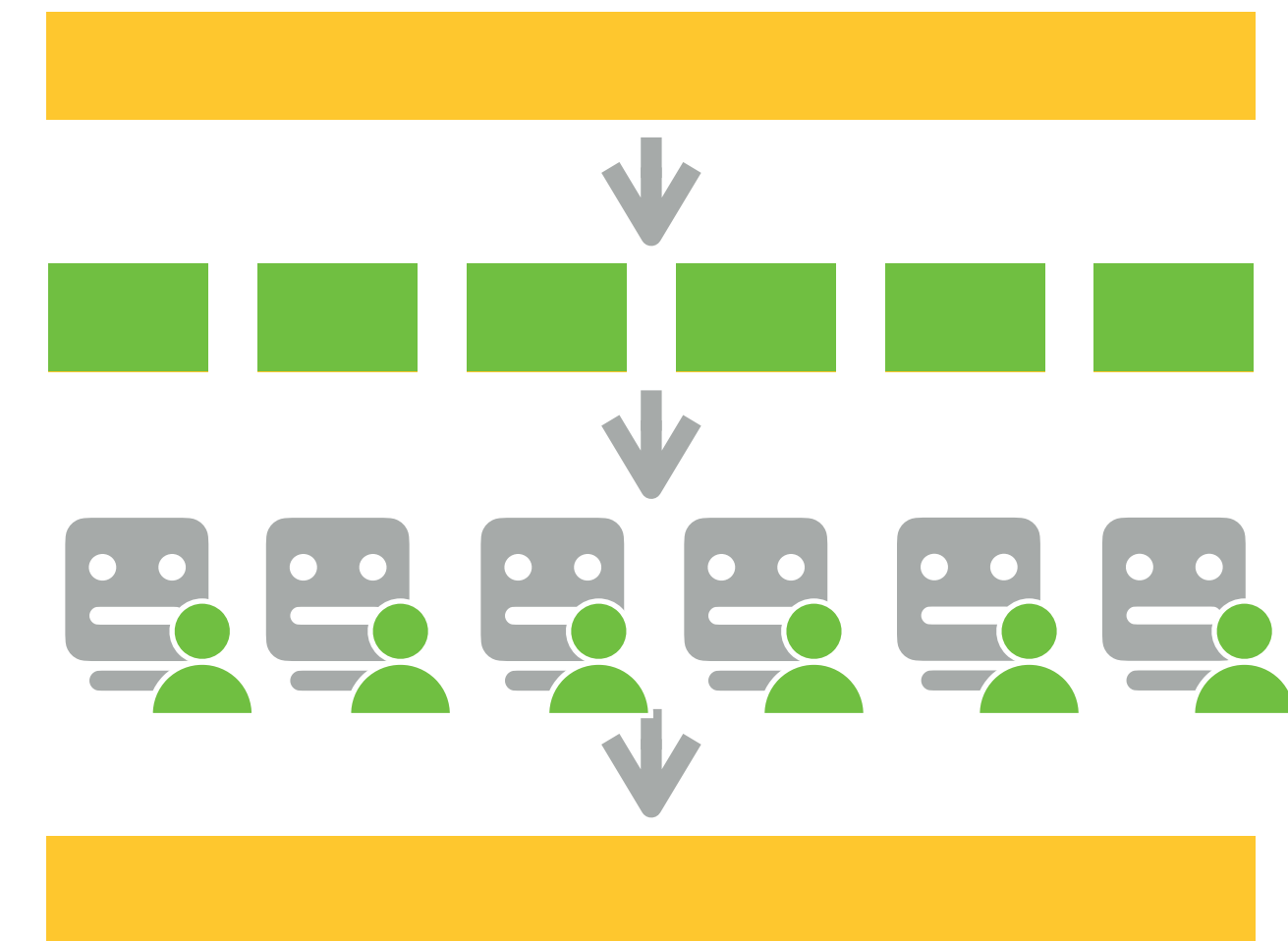
Limitations: cascading errors, conflicts between parallel paths, etc.

Crowdsourcing workflow vs. Chaining in LLMs

Crowdsourcing workflow



LLM chain



Differences...

Breakdown
rational

Address pitfalls of a single worker

Can do any task,
But do tasks with high variance,
Unwilling to digest too much context

Human
access

Humans only have access to the
steps assigned to them

More chance of conflict

Address pitfalls of a single LLM pass

Has intensive computing power,
But limited reasoning capability,
have exposure bias, etc.

Humans may interrupt at any step

Can take advantage of cascading effects,
parallel paths, for explainability

Human-AI Collaboration: key aspects

Complementary performance: Leverage the strengths of both AI and humans, to achieve better outcomes than either could accomplish alone.

Alignment: AI should behave in a ways that humans expect their “teammate” to behave.

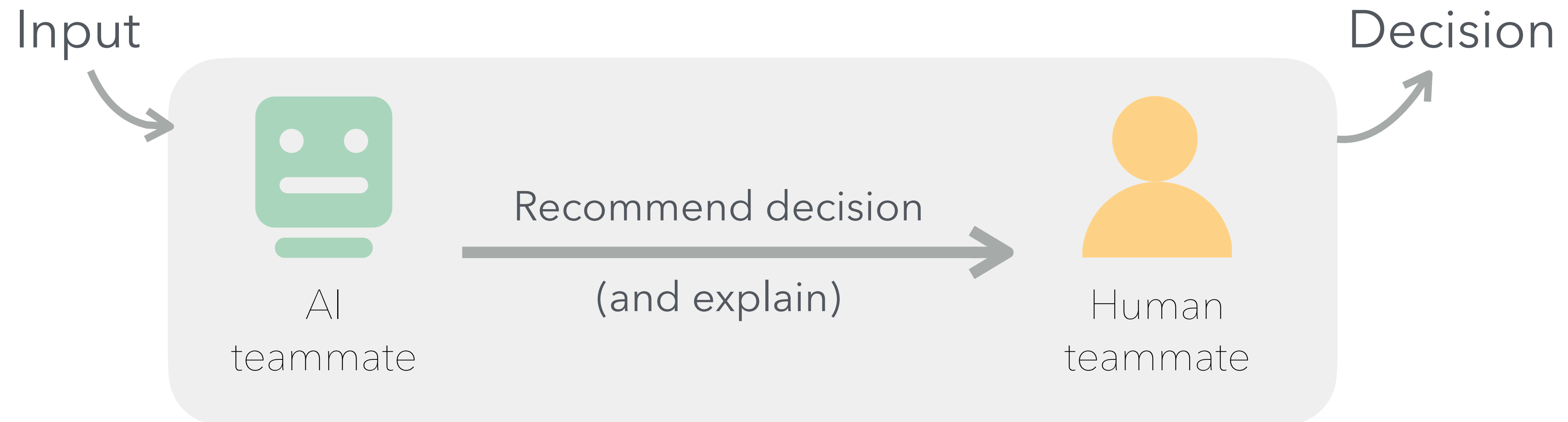
Cooperative interaction: The interaction between human and AI should work in a way that’s both intuitive for the human, and compatible to model characteristics.

Reasonable task allocation: Assign tasks that the model can perform.

Appropriate trust: humans should follow AIs when they are correct, but identify and correct their mistakes otherwise.

Appropriate trust in HAI team decision making

Human and AI perform the same task; AI makes suggested decisions and maybe explain, and human makes the final decision



Appropriate trust: *accept* AI recommendations when they are correct, but **overwrite** them when they don't make sense.

Design interactions for appropriate reliance

One example: When & how to display the AI recommendation matters.

inline highlight → confirmation bias ("this sentence seems reasonable enough")

See AI decision first → anchoring effect ("I will agree with AI's decision")

HCI solution

Display of AI recommendation

Present evidence, but not final decision; provide more explanations

Timing of AI decision

Asynchronous display, increase independence

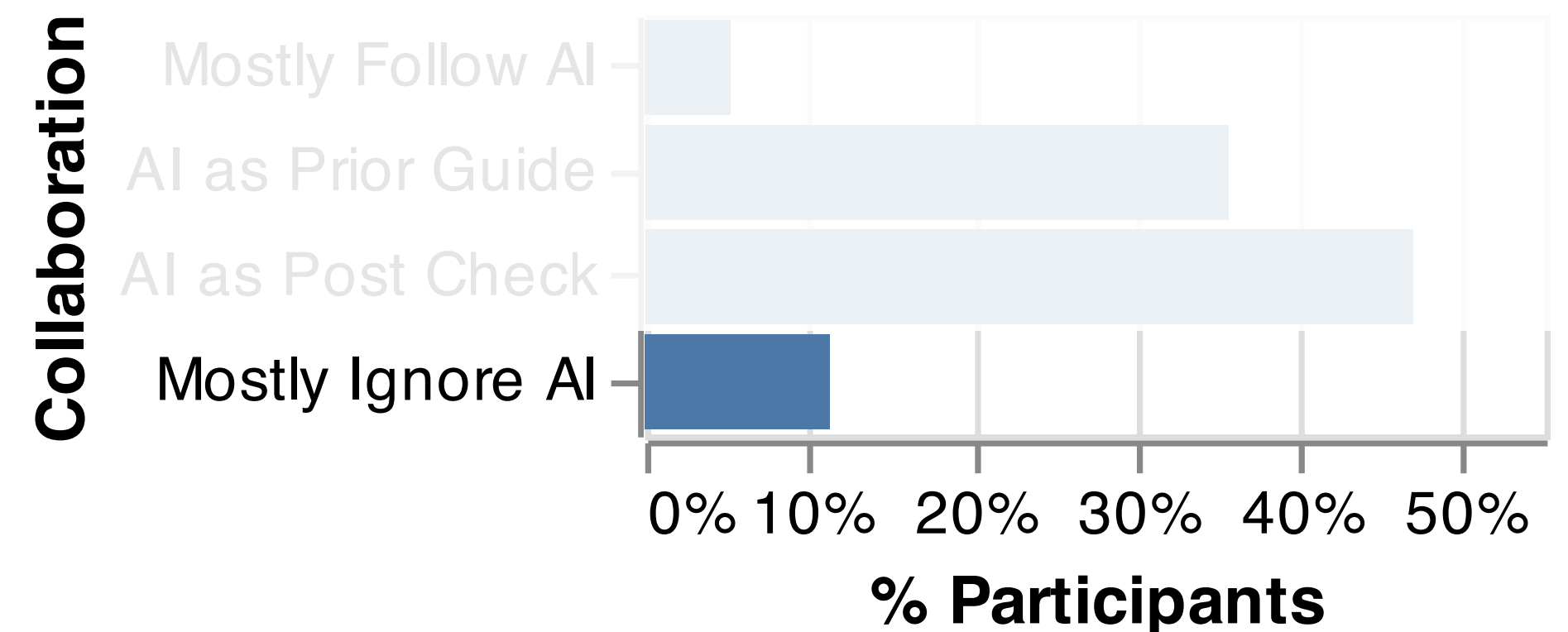
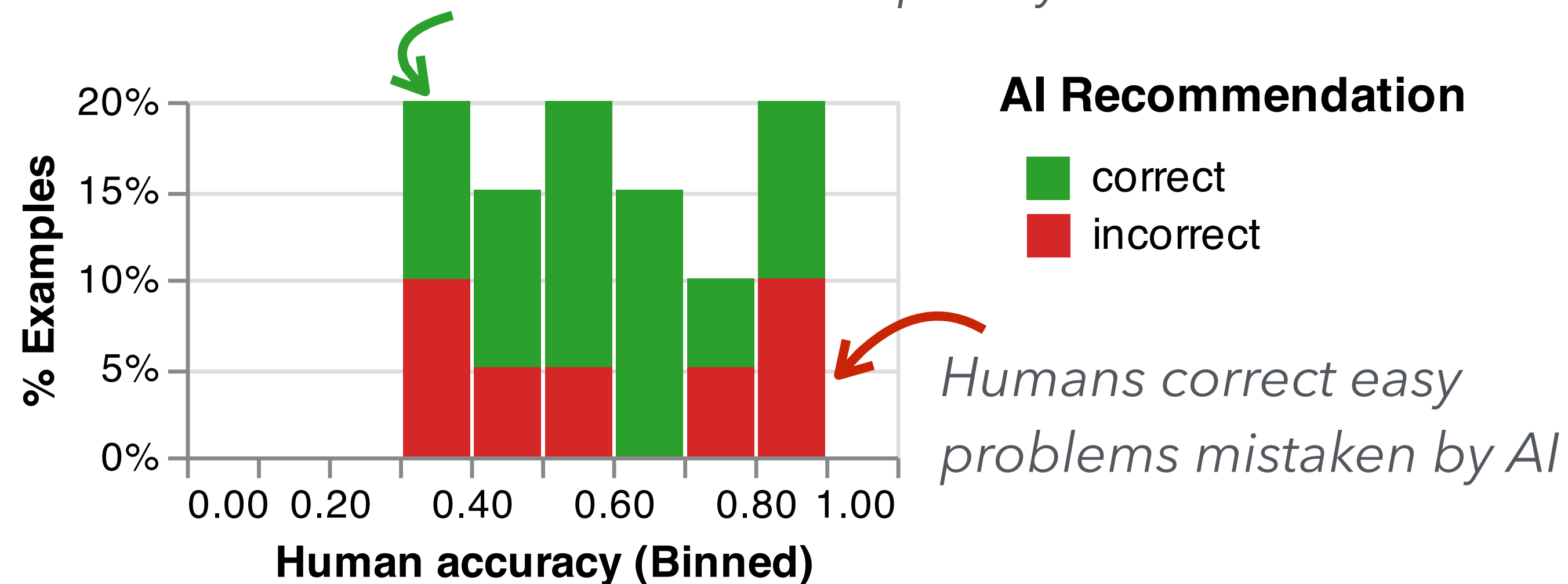
Factors causing inappropriate reliance?

The distribution of human and AI expertise matters.

Human+AI is ineffective if they make the same kinds of mistakes.

Multi-choice question answering task

AI adds value when humans frequently err

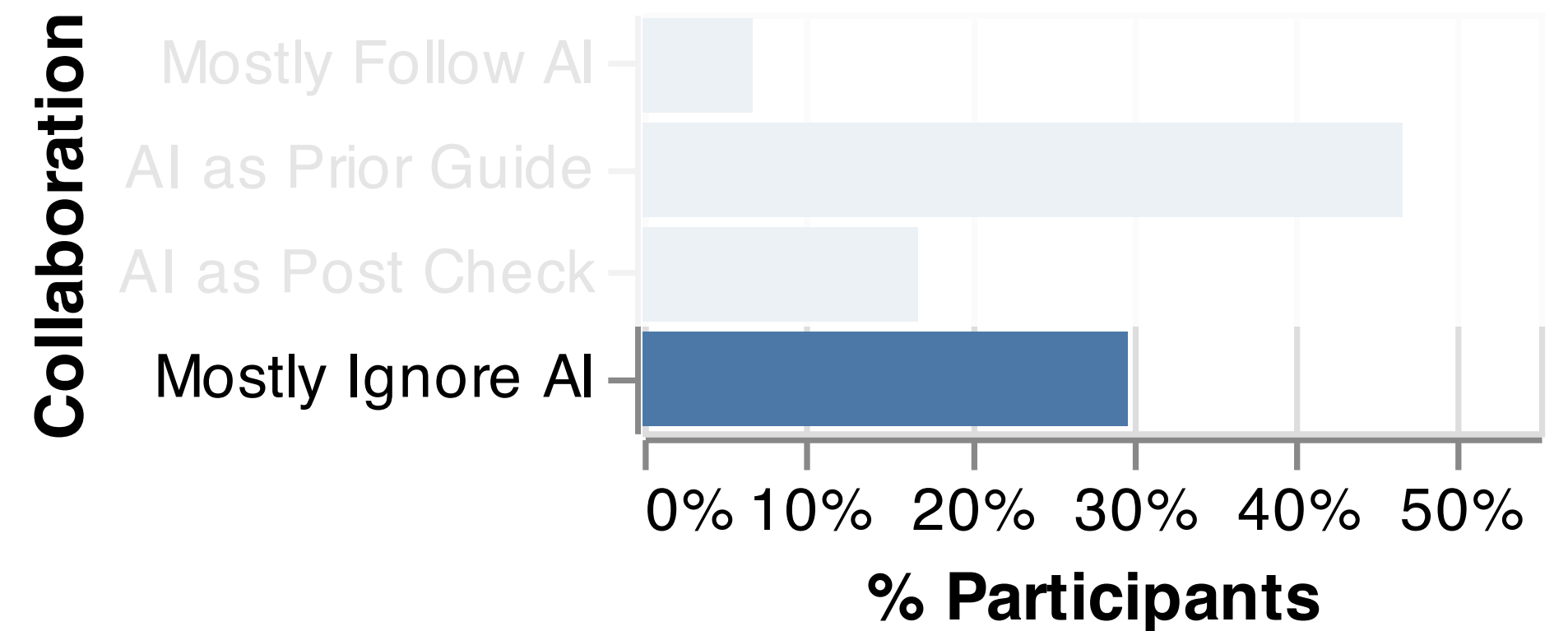
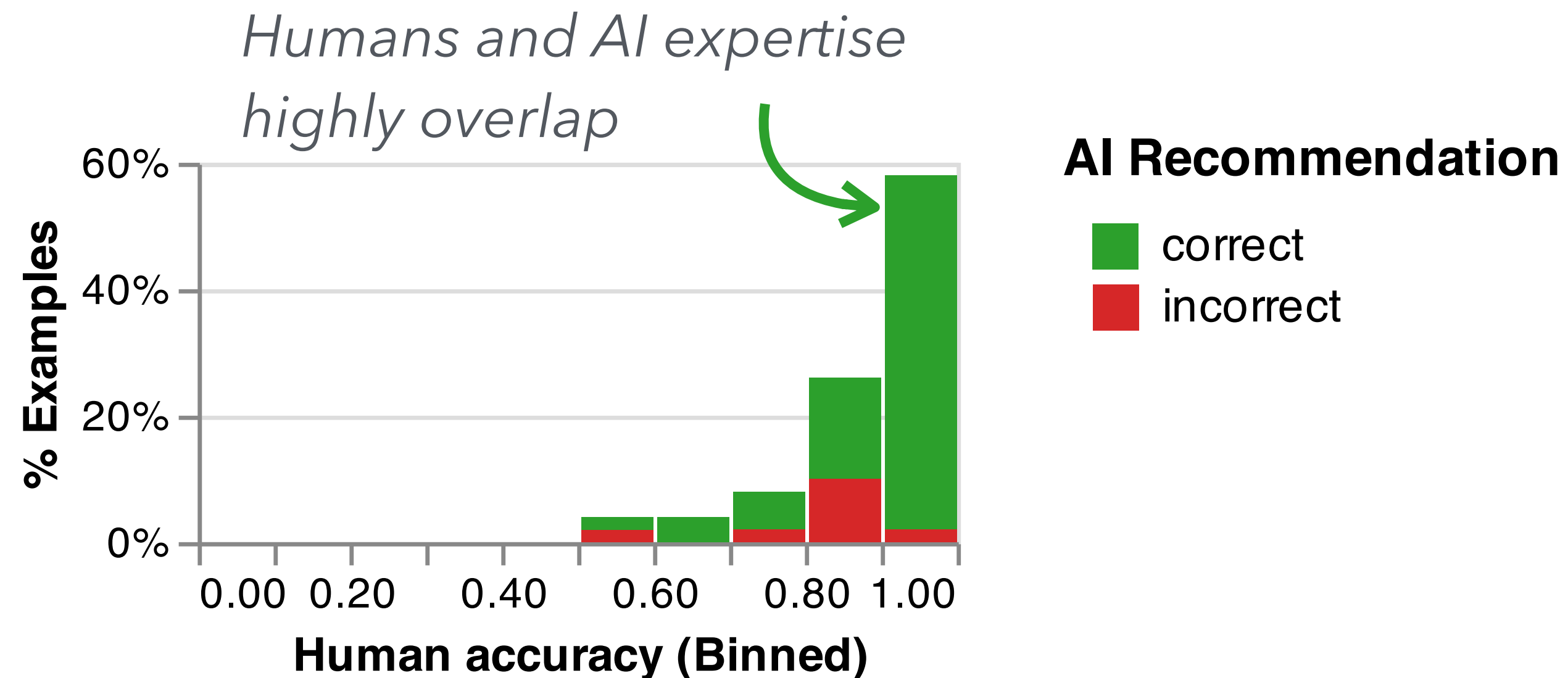


Factors causing inappropriate reliance?

The distribution of human and AI expertise matters.

Human+AI is ineffective if they make the same kinds of mistakes.

Classification task



Choose AIs carefully for the task

The distribution of human and AI expertise matters.

Human+AI is ineffective if they make the same kinds of mistakes.

AI solution

Change the training objective

Directly optimize for complementary behavior

HCI solution

Re-think AI's roles, to help in other dimensions

speed, cognitive load, etc.

Bansal, Gagan, et al. "Is the Most Accurate AI the Best Teammate? Optimizing AI for Teamwork." AAAI 2021
Feng, Shi, and Jordan Boyd-Graber. "What can ai do for me? evaluating machine learning interpretations in cooperative play." IUI 2019.
Glassman, Elena L., et al. "OverCode: Visualizing variation in student solutions to programming problems at scale." TOCHI 2015

Recap

Humans interact with models in different ways.

Key factors of Human-model interaction includes:

Complementary performance.

Alignment.

Reasonable task allocation.

Cooperative interaction.

Appropriate trust.

Prompting is cool, but needs to consider model capabilities.