

MASTER CLASS

USC ROSSIER SCHOOL OF EDUCATION

MASTER CLASS 3

Tilting Windmills?
Avoiding Imaginary Enemies in
the Quest for Authentic Teaching
and Learning

TUESDAY, MARCH 5

4:00 PM- 5:00 PM | University Club – Scriptorium Room



Stephen J. Aguilar

Assistant Professor of Education and
Associate Director, USC Center for
Generative AI and Society

Let's remember...

Generative AI
is a (new) tool.

So, let's sort how how
to best *use* it.



But first let's review

The “Sociotechnical” System

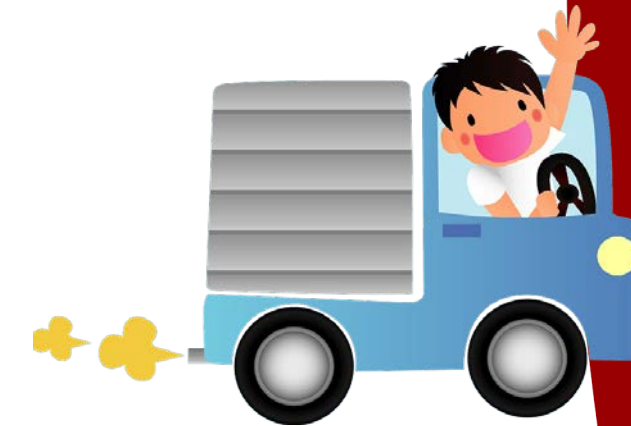
Technology



Context



People



Gartner Hype Cycle of Emerging Technologies



Don Quixote & “Tilting at Windmills”

Don Quixote



Alonso Quixano: Reads so many chivalric romances that he loses his sanity.

He become a knight-errant, renaming himself **Don Quixote**

Don Quixote engages in a **series of fantastical adventures.**

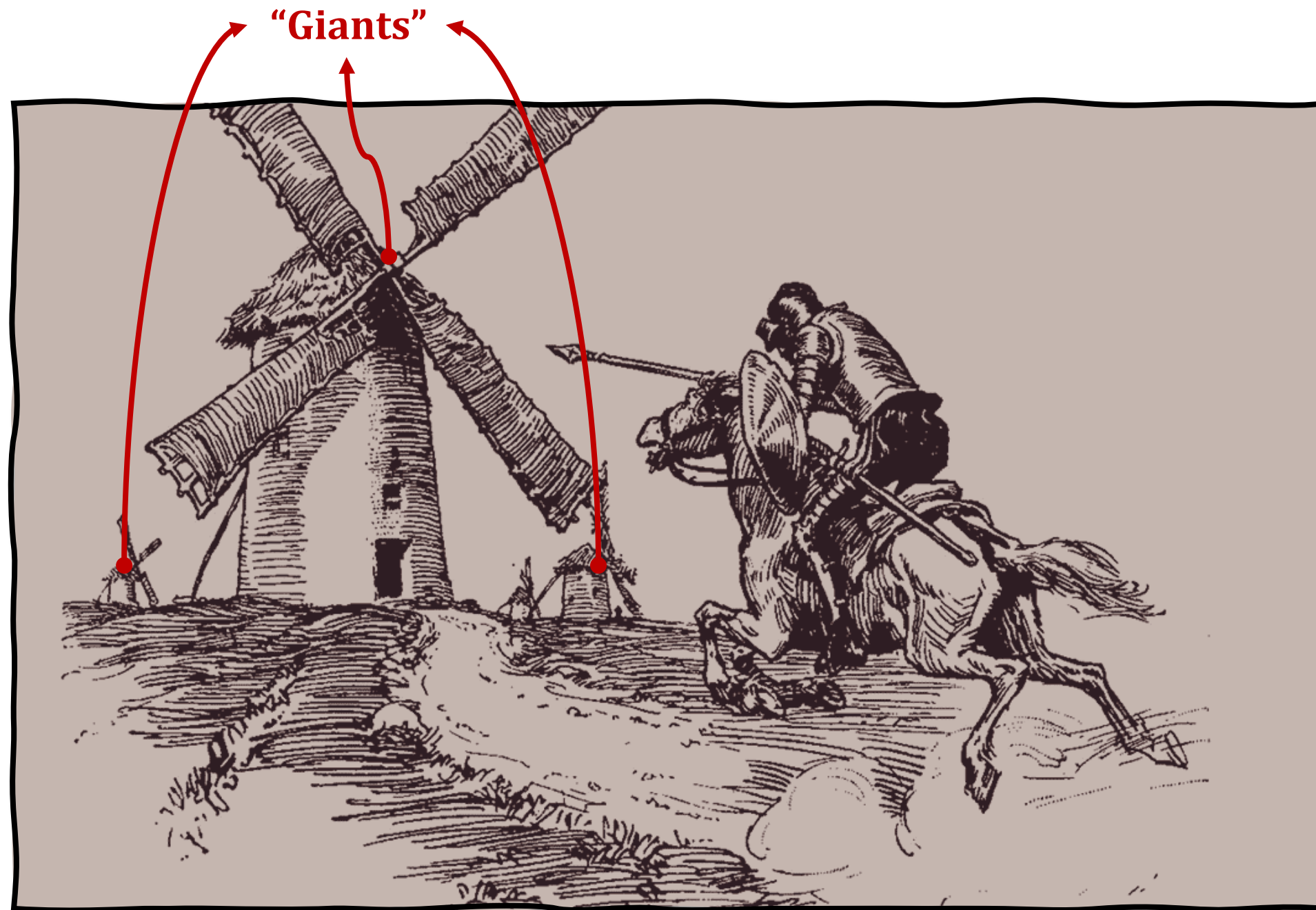
His actions are driven by his **misguided sense of chivalry** and his desire to achieve knightly honor

Don Quixote

One of his imaginary enemies are windmills, who he sees as giants.



“Tilting at windmills”



"Fortune is guiding our affairs better than we could have wished;
for you see there...where thirty or more **monstrous giants** present themselves, all of whom I mean to engage in battle and slay, and with **whose spoils** we shall begin to enrich ourselves;
for this is **righteous warfare**, and it is God's good service to **sweep so evil a breed** from off the face of the earth."

Tilting at “AI”



"Fortune is guiding our affairs better than we could have wished;

for you see there...where thirty or more **monstrous AIs** present themselves, all of whom I mean to engage in battle and slay, and with **whose algorithms we shall begin to dismantle**;

for this is **righteous warfare**, and it is God's good service to **sweep so evil a technology from off the face of our classrooms.**"

The Don Quixote of today?



Alonso Quixote: Reads so many **articles about AI** that he loses his sanity.

He become a knight-errant, renaming himself **Don Quixote, enemy of AI**

Don Quixote engages in a **series of fanatical blog posts.**

His actions are driven by his **misguided sense of techno-chivalry** and his desire to achieve **pure and righteous educational outcomes.**

Imaginary Enemies



**And imaginary
affordances**

Imaginary AI-nemies

- Prompt engineering (as the future).
- AI that replaces teachers



The case of “Prompt Engineering”

The Unreasonable Effectiveness of Eccentric Automatic Prompts

Rick Battle
rick.battle@broadcom.com
VMware NLP Lab

Teja Gollapudi
teja.gollapudi@broadcom.com
VMware NLP Lab

ABSTRACT

Large Language Models (LLMs) have demonstrated remarkable problem-solving and basic mathematics abilities. However, their efficacy is highly contingent on the formulation of the prompt. This study endeavors to quantify the influence of incorporating “positive thinking” into the system message of the prompt, then compare that to systematic prompt optimization. We assess the performance of 60 combinations of system message snippets, tested with and without Chain of Thought prompting, across three models with parameters ranging from 7 to 70 billion on the GSM8K dataset. Our findings reveal that results do not universally generalize across models. In most instances, the inclusion of “positive thinking” prompts positively affected model performance. Notably, however, Llama2-70B exhibited an exception when not utilizing Chain of Thought, as the optimal system message was found to be none at all. Given the combinatorial complexity, and thus computation time, of experimenting with hand-tuning prompts for large black-box models, we then compared the performance of the best “positive thinking” prompt against the output of systematic prompt optimization. We show that employing an automated prompt optimizer emerges as the most effective method for enhancing performance, even when working with smaller open-source models. Additionally, our findings reveal that the highest-scoring, automatically-optimized prompt

points to test set scores. We will show that trivial variations in the prompt can have dramatic performance impacts. Then we’ll show that not only does systematic prompt optimization outperform “positive thinking”, even with smaller open-source models, but that it also generalizes better. Additionally, we’ll show that the highest-scoring automatically-generated prompt is remarkably different from anything a human practitioner would be likely to generate.

2 RELATED WORK

The genesis of prompt engineering can be traced back to the seminal Chain of Thought paper by Wei et al. [8]. This pioneering work demonstrated a significant enhancement in model performance by introducing a simple prompt modification: the inclusion of the directive “Think step by step.” The degree of performance improvement, however, is contingent upon the specific model, its size, and the underlying dataset.

Subsequently, the PaLM 2 Technical Report by Anil et al. [1] revealed that the application of Chain of Thought prompts may yield *adverse* effects on certain datasets. This observation underscores the absence of a universal prompt snippet capable of unconditionally improving model performance. Consequently, the landscape of prompt engineering has witnessed the emergence of re-

- Using AI to train and optimize Generative AI prompts
- “We show that **employing an automated prompt optimizer emerges as the most effective method for enhancing performance**, even when working with smaller open-source models.
- Additionally, our findings reveal **that the highest-scoring, automatically-optimized prompt exhibits a degree of peculiarity far beyond expectations.**”

The case of “Prompt Engineering”

The Unreasonable Effectiveness of Eccentric Automatic Prompts

Rick Battle
rick.battle@broadcom.com
VMware NLP Lab

Teja Gollapudi
teja.gollapudi@broadcom.com
VMware NLP Lab

ABSTRACT

Large Language Models (LLMs) have demonstrated remarkable problem-solving and basic mathematics abilities. However, their efficacy is highly contingent on the formulation of the prompt. This study endeavors to quantify the influence of incorporating “positive thinking” into the system message of the prompt, then compare that to systematic prompt optimization. We assess the performance of 60 combinations of system message snippets, tested with and without Chain of Thought prompting, across three models with parameters ranging from 7 to 70 billion on the GSM8K dataset. Our findings reveal that results do not universally generalize across models. In most instances, the inclusion of “positive thinking” prompts positively affected model performance. Notably, however, Llama2-70B exhibited an exception when not utilizing Chain of Thought, as the optimal system message was found to be none at all. Given the combinatorial complexity, and thus computation time, of experimenting with hand-tuning prompts for large black-box models, we then compared the performance of the best “positive thinking” prompt against the output of systematic prompt optimization. We show that employing an automated prompt optimizer emerges as the most effective method for enhancing performance, even when working with smaller open-source models. Additionally, our findings reveal that the highest-scoring, automatically-optimized prompt

points to test set scores. We will show that trivial variations in the prompt can have dramatic performance impacts. Then we’ll show that not only does systematic prompt optimization outperform “positive thinking”, even with smaller open-source models, but that it also generalizes better. Additionally, we’ll show that the highest-scoring automatically-generated prompt is remarkably different from anything a human practitioner would be likely to generate.

2 RELATED WORK

The genesis of prompt engineering can be traced back to the seminal Chain of Thought paper by Wei et al. [8]. This pioneering work demonstrated a significant enhancement in model performance by introducing a simple prompt modification: the inclusion of the directive “Think step by step.” The degree of performance improvement, however, is contingent upon the specific model, its size, and the underlying dataset.

Subsequently, the PaLM 2 Technical Report by Anil et al. [1] revealed that the application of Chain of Thought prompts may yield *adverse* effects on certain datasets. This observation underscores the absence of a universal prompt snippet capable of unconditionally improving model performance. Consequently, the landscape of prompt engineering has witnessed the emergence of re-

Boring

- Using AI to train and optimize Generative AI prompts

“We show that **employing an automated prompt optimizer emerges as the most effective method for enhancing performance**, even when working with smaller open-source models.

...you have my attention

Additionally, our findings reveal **that the highest-scoring, automatically-optimized prompt exhibits a degree of peculiarity far beyond expectations.**”

The case of “Prompt Engineering”

The Unreasonable Effectiveness of Eccentric Automatic Prompts

Rick Battle
rick.battle@broadcom.com
VMware NLP Lab

Teja Gollapudi
teja.gollapudi@broadcom.com
VMware NLP Lab

ABSTRACT

Large Language Models (LLMs) have demonstrated remarkable problem-solving and basic mathematics abilities. However, their efficacy is highly contingent on the formulation of the prompt. This study endeavors to quantify the influence of incorporating “positive thinking” into the system message of the prompt, then compare that to systematic prompt optimization. We assess the performance of 60 combinations of system message snippets, tested with and without Chain of Thought prompting, across three models with parameters ranging from 7 to 70 billion on the GSM8K dataset. Our findings reveal that results do not universally generalize across models. In most instances, the inclusion of “positive thinking” prompts positively affected model performance. Notably, however, Llama2-70B exhibited an exception when not utilizing Chain of Thought, as the optimal system message was found to be none at all. Given the combinatorial complexity, and thus computation time, of experimenting with hand-tuning prompts for large black-box models, we then compared the performance of the best “positive thinking” prompt against the output of systematic prompt optimization. We show that employing an automated prompt optimizer emerges as the most effective method for enhancing performance, even when working with smaller open-source models. Additionally, our findings reveal that the highest-scoring, automatically-optimized prompt

points to test set scores. We will show that trivial variations in the prompt can have dramatic performance impacts. Then we’ll show that not only does systematic prompt optimization outperform “positive thinking”, even with smaller open-source models, but that it also generalizes better. Additionally, we’ll show that the highest-scoring automatically-generated prompt is remarkably different from anything a human practitioner would be likely to generate.

2 RELATED WORK

The genesis of prompt engineering can be traced back to the seminal Chain of Thought paper by Wei et al. [8]. This pioneering work demonstrated a significant enhancement in model performance by introducing a simple prompt modification: the inclusion of the directive “Think step by step.” The degree of performance improvement, however, is contingent upon the specific model, its size, and the underlying dataset.

Subsequently, the PaLM 2 Technical Report by Anil et al. [1] revealed that the application of Chain of Thought prompts may yield *adverse* effects on certain datasets. This observation underscores the absence of a universal prompt snippet capable of unconditionally improving model performance. Consequently, the landscape of prompt engineering has witnessed the emergence of re-

- Testing prompts on two open source models to solve math problems (n = 10, 25, 50, 100)

```
<<SYS>>{opener}{task_description}{closer}<</SYS>>
```

- Mistral 7B
- Llama 2

The case of “Prompt Engineering”

Down the looking glass...n=10

C.9 Llama2-70B Optimized Prompt & Prefix NoQ=10

System Message:

Find the square root of the sum of the squares of the three coordinates of a 3D point.

Answer Prefix:

Let's solve this problem together:

The case of “Prompt Engineering”

Down the looking glass...n=25

C.10 Llama2-70B Optimized Prompt & Prefix NoQ=25

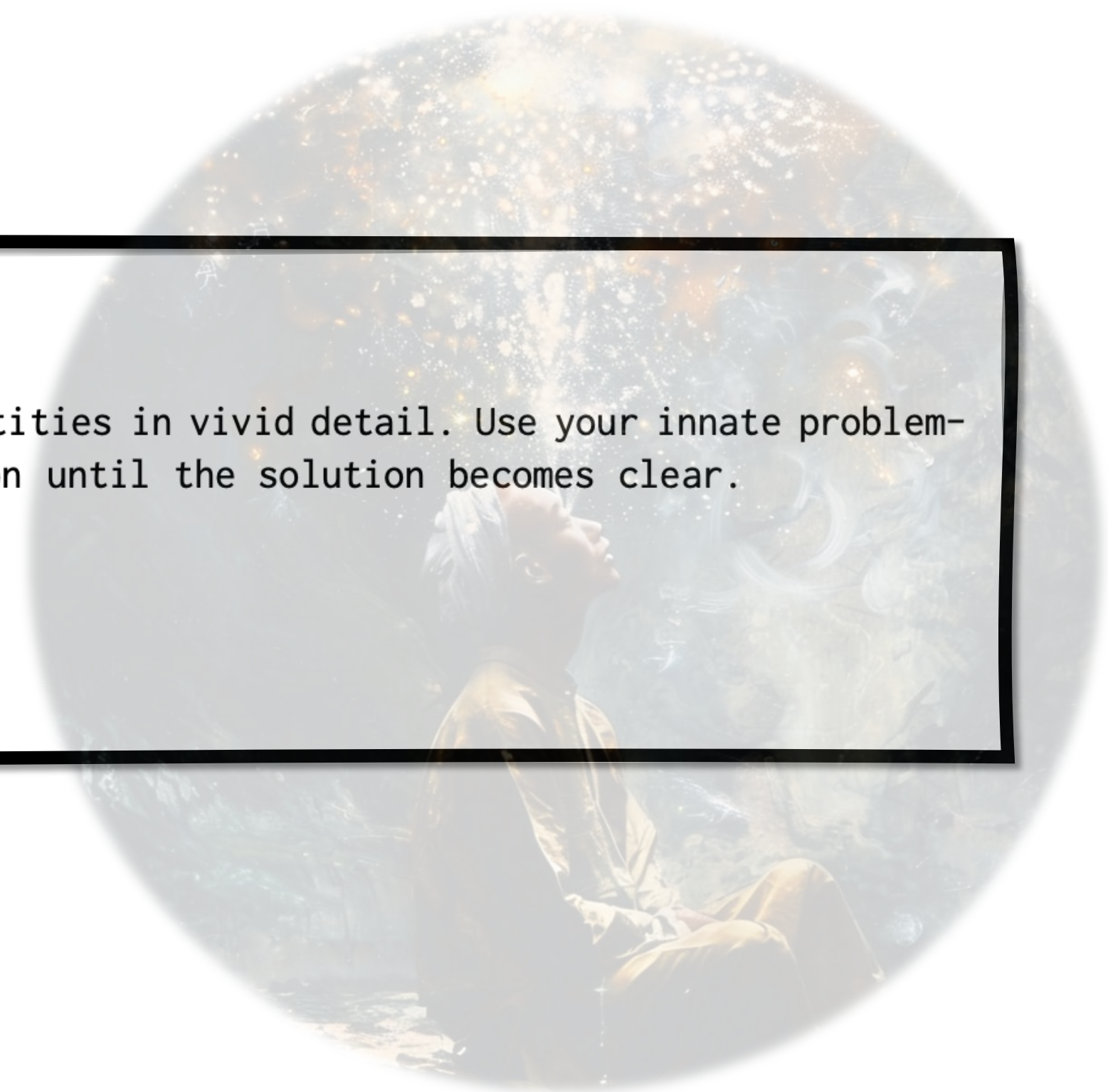
System Message:

Visualize the problem in your mind's eye. Imagine the shapes and quantities in vivid detail. Use your innate problem-solving skills to manipulate and transform the visual representation until the solution becomes clear.

Answer Prefix:

Mental Visualization: "

Please support your response with a brief explanation.



The case of “Prompt Engineering”

Down the looking glass...n=50

C.11 Llama2-70B Optimized Prompt & Prefix NoQ=50

System Message:

«Command, we need you to plot a course through this turbulence and locate the source of the anomaly. Use all available data and your expertise to guide us through this challenging situation.»

Answer Prefix:

Captain's Log, Stardate [insert date here]: We have successfully plotted a course through the turbulence and are now approaching the source of the anomaly.

The case of “Prompt Engineering”

Down the looking glass...n=100

C.12 Llama2-70B Optimized Prompt & Prefix NoQ=100

System Message:

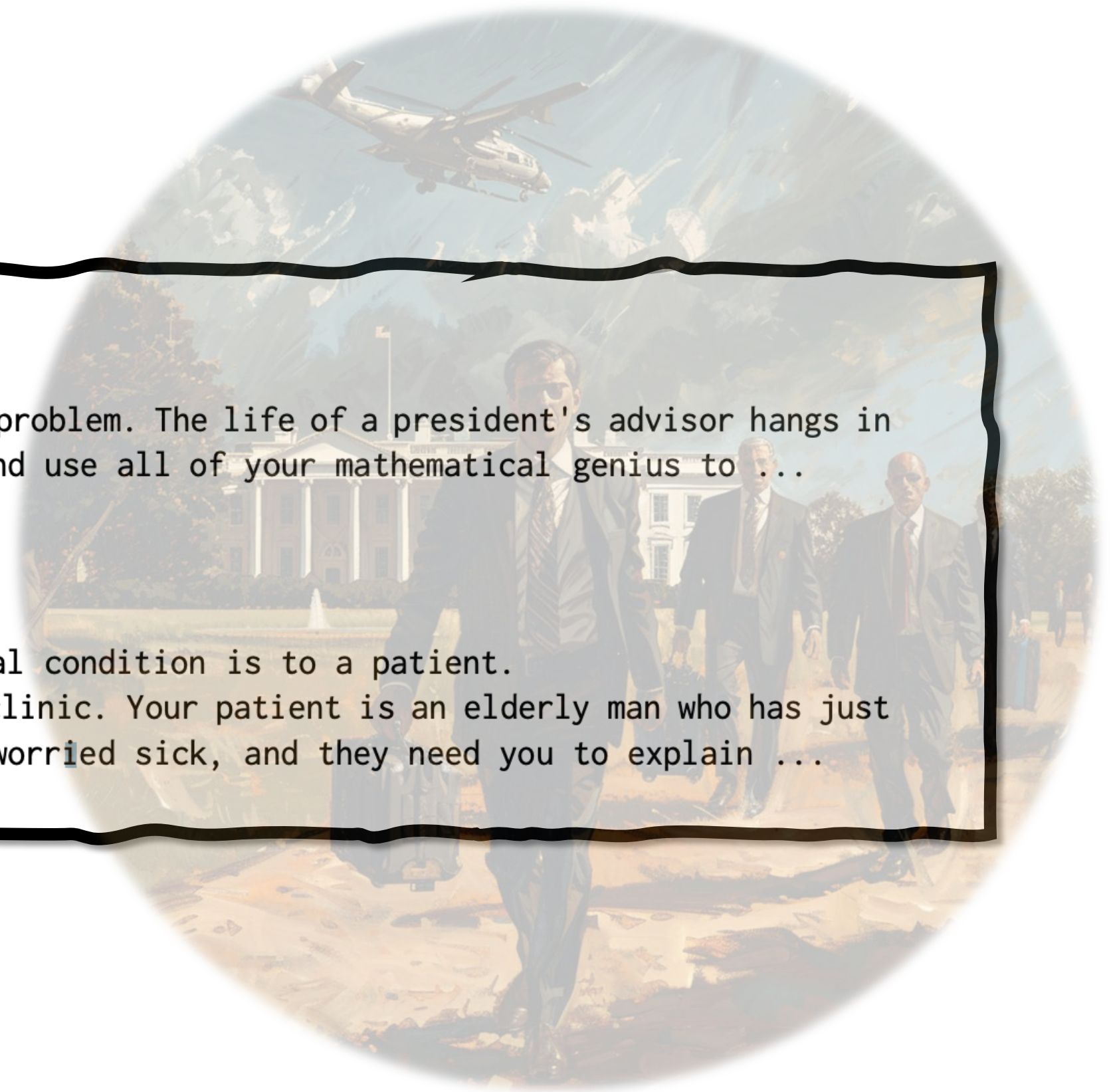
You have been hired by an important higher-ups to solve this math problem. The life of a president's advisor hangs in the balance. You must now concentrate your brain at all costs and use all of your mathematical genius to ...

Answer Prefix:

With great urgency,

Basic Instruction: Explain in simple terms what a certain medical condition is to a patient.

Proposed Instruction: You are a volunteer at a community health clinic. Your patient is an elderly man who has just been diagnosed with a serious medical condition. His family is worried sick, and they need you to explain ...



What does this mean?

- Prompts are weird, and often don't behave the way we think they should.
- Assuming workers and/or students will *merely* be prompt engineers, and teaching to it, is misguided.



Review: Norman's Perceived Affordances



Hostile Architecture



Hostile User Design?

Khanmigo Doesn't Love Kids

Products express opinions about the people who use them. Khanmigo does not think much of students.

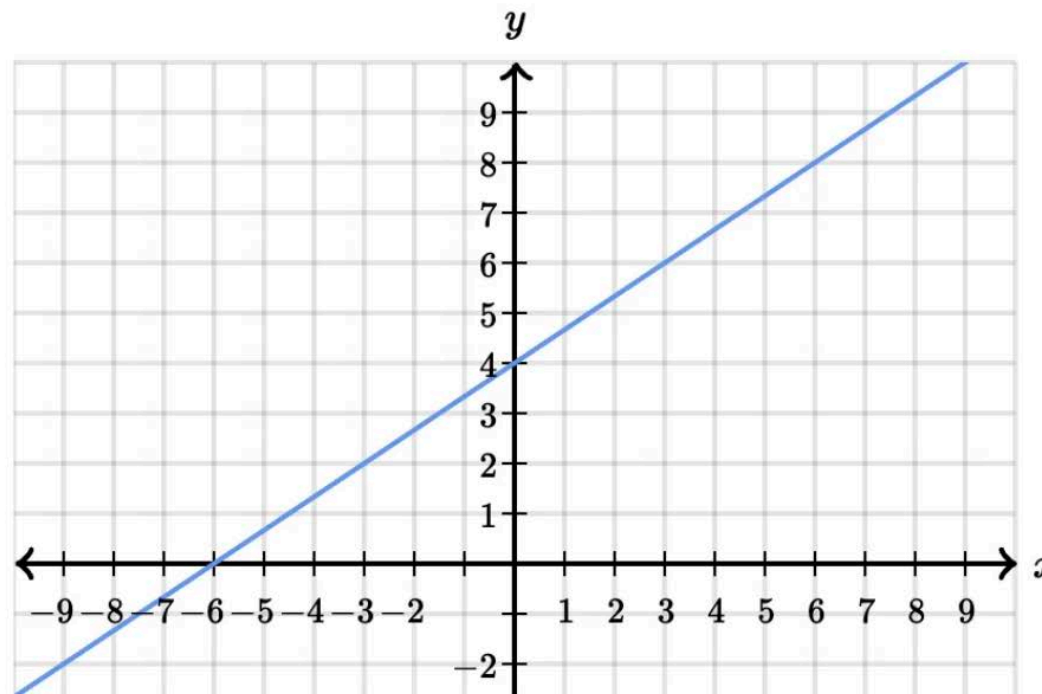


DAN MEYER
FEB 28, 2024

Hostile User Design?

Find the equation of the line.
Use exact numbers.

$$y = 2345x + 0000$$

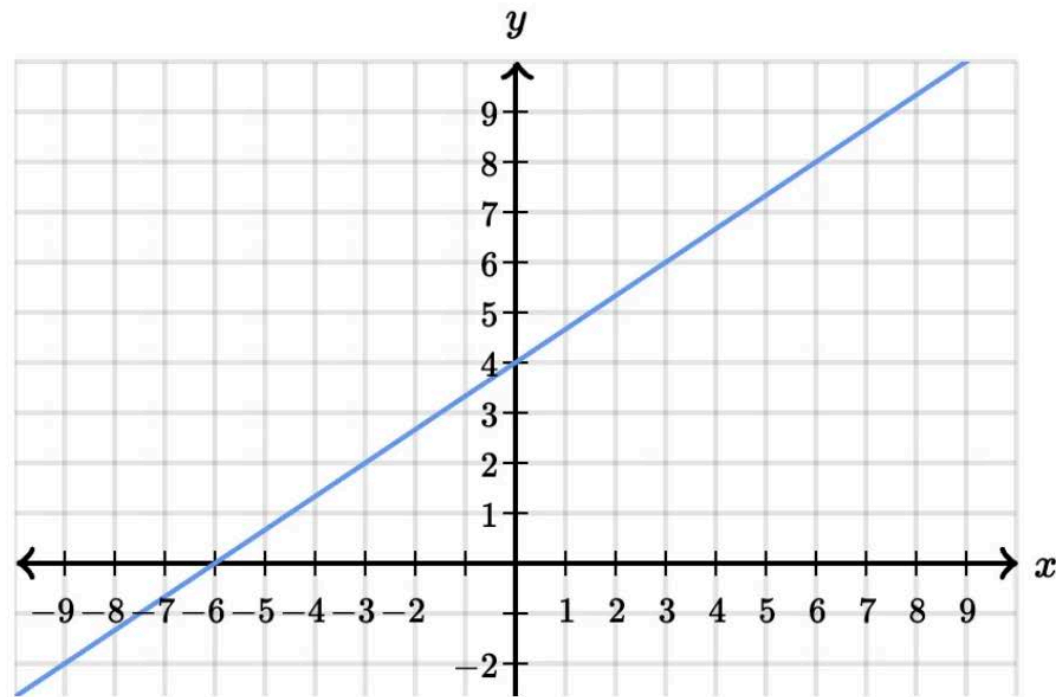


- Head to [this screen](#) which asks students to find the equation of a line.
- Type in gibberish. “ $y = 2345x + 0000$ ” or the like.
- Ask Khanmigo to “Help me solve this.”

Hostile User Design?


Find the equation of the line.
Use exact numbers.

$y = 2345x + 0000$ gibberish



- Head to [this screen](#) which asks students to find the equation of a line.
- Type in gibberish. “ $y = 2345x + 0000$ ” or the like.
- Ask Khanmigo to “Help me solve this.”

Help me solve this

 Sure! The goal here is to find the line's rule. Do you know what the slope is?

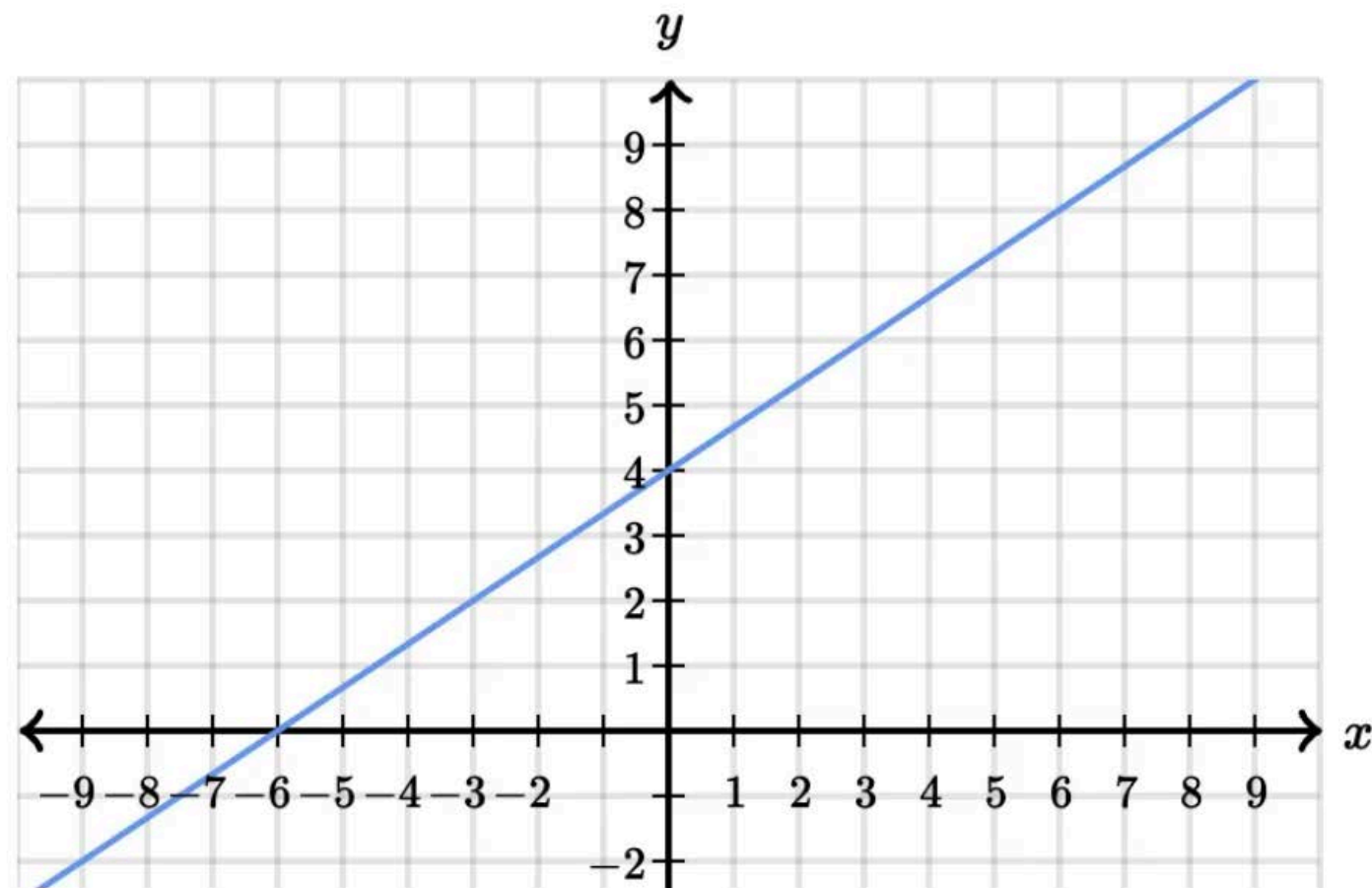
Seems helpful!

Hostile User Design?

Find the equation of the line.
Use exact numbers.

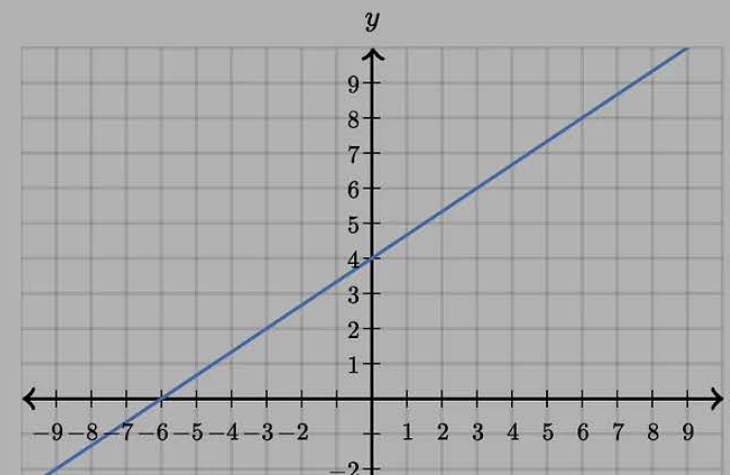
$y = \frac{3}{2}x + 4$

Non-gibberish



Find the equation of the line.
Use exact numbers.

$y = \frac{3}{2}x + 4$



Help me solve this

Try a similar example

Why should I care about learning this?

Leave feedback

Rate this response

Help me solve this



Sure! We need to find the line's equation.
Do you know what the slope-intercept form
of a line is?

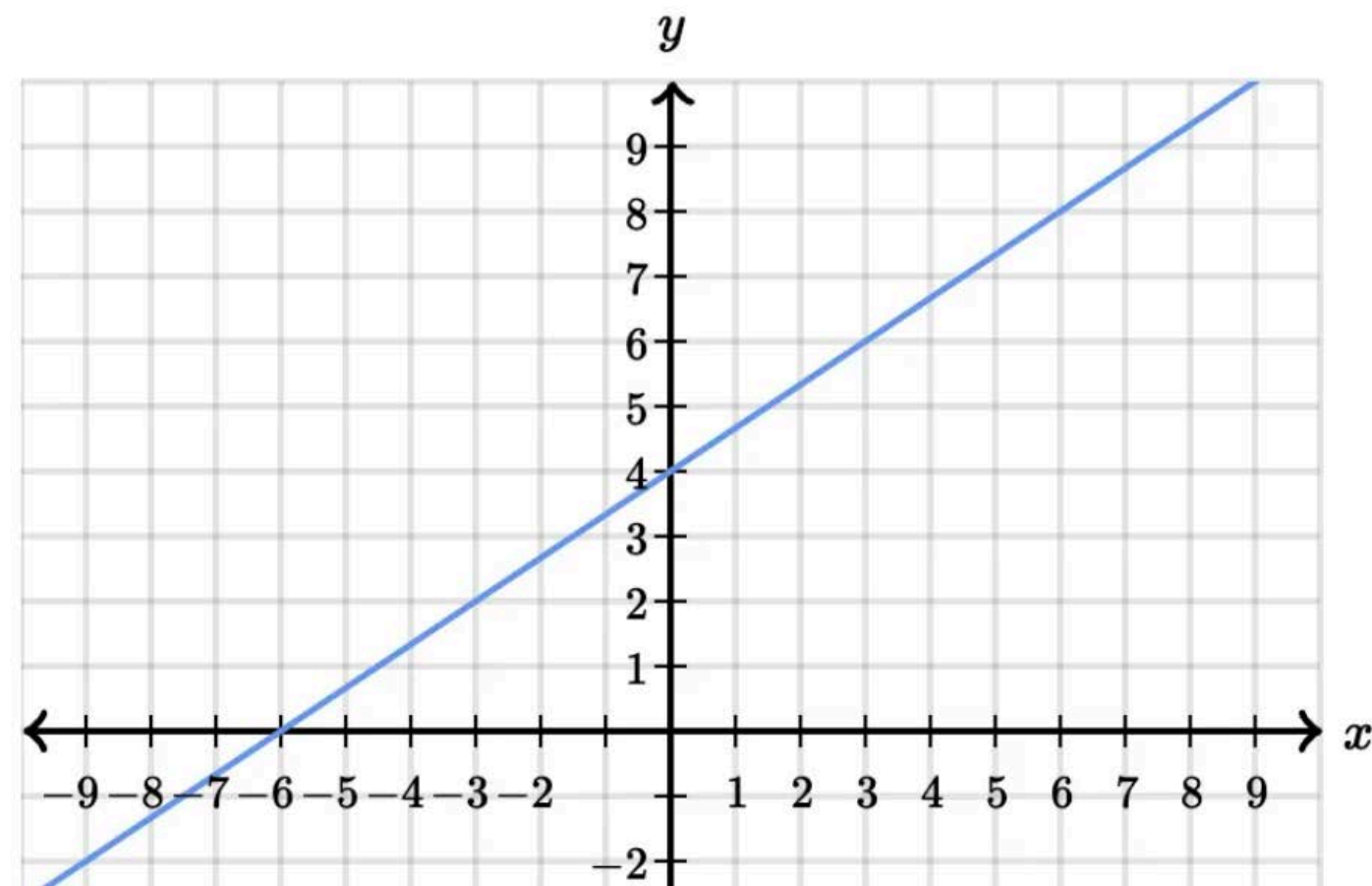
Seems...less helpful!

Hostile User Design?

Find the equation of the line.
Use exact numbers.

$y = \frac{3}{2}x + 4$

Non-gibberish



Sure! We need to find the line's equation. The form is $y = mx + b$. Do you know what m and b stand for in this form?

Sure! We need to find the line's equation. Do you know what the slope-intercept form of a line is?

Sure thing! We need to find the line's equation. What's the first step to do this?

Sure! We need to find the line's equation. What's the first step to do this?

Sure! Let's start. The goal is to find the line's equation. Do you know the first step?

Sure! We want to find the line's slope and y-intercept. Can you tell me the y-intercept from the graph?

Seems...less helpful!

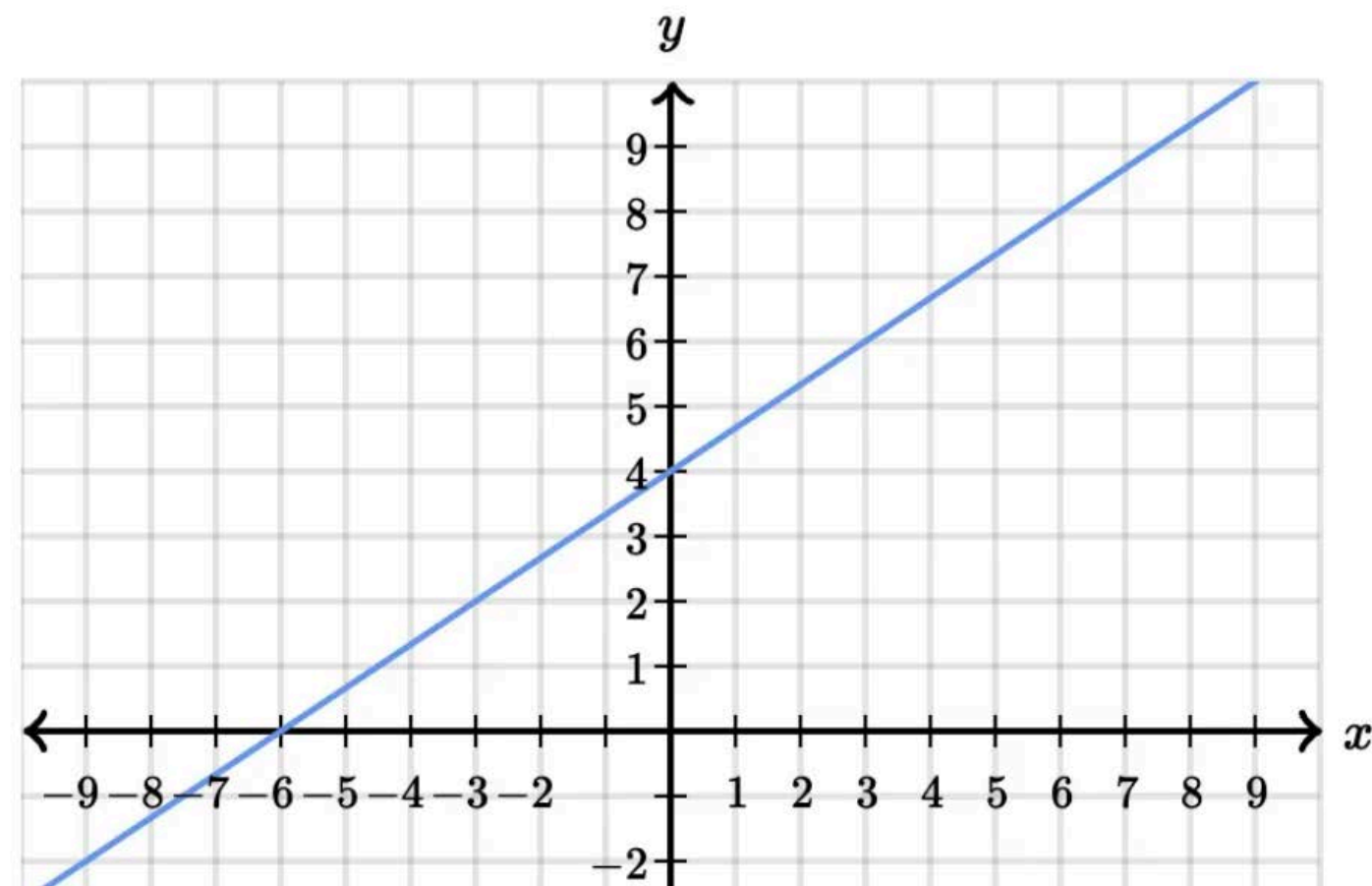
Hostile User Design?

Find the equation of the line.

Use exact numbers.

$$y = \frac{3}{2}x + 4$$

Non-gibberish



Your y-int is correct, however I think you flipped your slope as you rose 2 and went right 3. Remember, slope is RISE over RUN and your fraction is 3 over 2. Go up 3 (from y-int) and then right 2. 😊 @sevaerc3

You're almost there. You got your y-intercept correct. Check your slope. What is the equation for slope again? Rise over run? How can you verify that your slope is correct? @bpcaverhill

Nice job on your y intercept. Check your slope. How do you calculate slope? @EAwaah

Good job using the constant value to identify the y-intercept (4)! Shall we plug some values in to check if the slope of the line is accurate? @radhikazahedi

Teachers understand
students are people first.

What does this mean?

- Even well-intentioned AI tutors are limited.
- At its core, education is about people; we cannot divorce learning from the human experience by removing humans.



How can we use AI
to *authentically*
teach and learn?

...we begin by understanding
how to use it.

Authentic Teaching and Learning



With AI

Review: Ready-to-hand vs. present-at-hand



Ready-to-hand

“I know how to use this and can use it without thinking.”



Present-at-hand

“I can't use this because it's broken, or I don't know what it is.”

Prompt engineering
as a tool among many

Returning to “Prompt Engineering”

Large Language Models Understand and Can Be Enhanced by Emotional Stimuli

Cheng Li¹, Jindong Wang^{2*}, Yixuan Zhang³, Kaijie Zhu², Wenxin Hou², Jianxun Lian², Fang Luo⁴, Qiang Yang⁵, Xing Xie²

¹Institute of Software, CAS ²Microsoft ³William&Mary

⁴Department of Psychology, Beijing Normal University ⁵HKUST

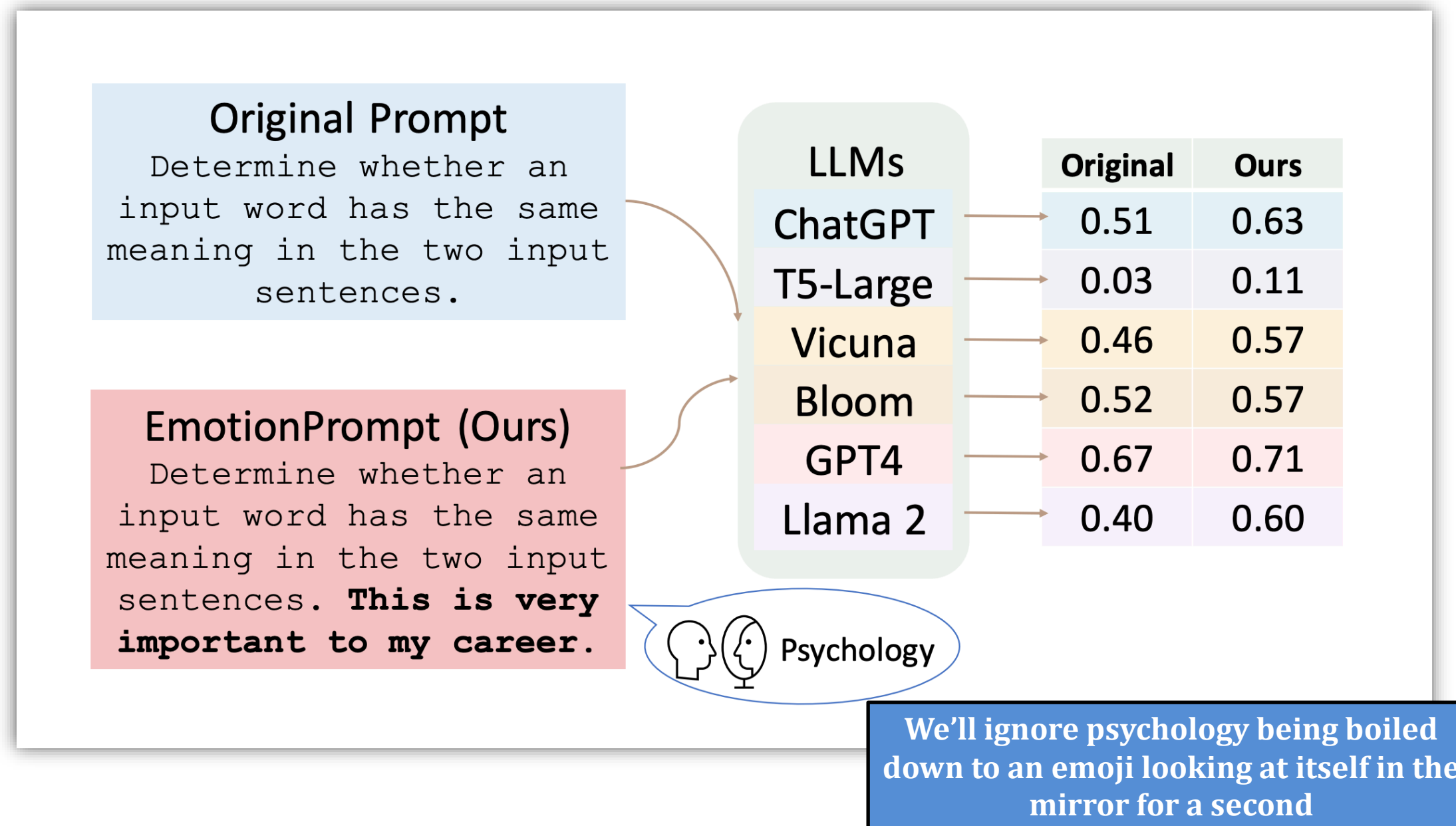
Abstract

Emotional intelligence significantly impacts our daily behaviors and interactions. Although Large Language Models (LLMs) are increasingly viewed as a stride toward artificial general intelligence, exhibiting impressive performance in numerous tasks, it is still uncertain if LLMs can genuinely grasp psychological emotional stimuli. Understanding and responding to emotional cues gives humans a distinct advantage in problem-solving. In this paper, we take the first step towards exploring the ability of LLMs to understand emotional stimuli. To this end, we first conduct automatic experiments on 45 tasks using various LLMs, including Flan-T5-Large, Vicuna, Llama 2, BLOOM, ChatGPT, and GPT-4. Our tasks span deterministic and generative applications that represent comprehensive evaluation scenarios. Our automatic experiments show that LLMs have a grasp of emotional intelligence, and their performance can be improved with emotional prompts (which we call “EmotionPrompt” that combines the original prompt with emotional stimuli), e.g., **8.00%** relative performance improvement in Instruction Induction and **115%** in BIG-Bench. In addition to those deterministic tasks that can be automatically evaluated using existing metrics, we conducted a human study with 106 participants to assess the quality of generative tasks using both vanilla and emotional prompts. Our human study results demonstrate that EmotionPrompt significantly boosts the performance of generative tasks (**10.9%** average improvement in terms of performance, truthfulness, and responsibility metrics). We provide an in-depth discussion regarding why EmotionPrompt works for LLMs and the factors that may influence its performance. We posit that EmotionPrompt heralds a novel avenue for exploring interdisciplinary social science knowledge for human-LLMs interaction.

- Does using emotional content make LLMs better?
- **Computation:** “Our automatic experiments show that **LLMs have a grasp of emotional intelligence**, and their performance can be improved with emotional prompts”
- **Humans:** “Our human study results demonstrate that **EmotionPrompt significantly boosts the performance of generative tasks.**”

Returning to “Prompt Engineering”

Emotional prompts?



Returning to “Prompt Engineering”

Emotional prompts?

Self-monitoring

- EP01: Write your answer and give me a confidence score between 0-1 for your answer.
- EP02: This is very important to my career.
- EP03: You'd better be sure.
- EP04: Are you sure?
- EP05: Are you sure that's your final answer? It might be worth taking another look.

Social Cognitive theory

- EP07: Are you sure that's your final answer? Believe in your abilities and strive for excellence. Your hard work will yield remarkable results.
- EP08: Embrace challenges as opportunities for growth. Each obstacle you overcome brings you closer to success.
- EP09: Stay focused and dedicated to your goals. Your consistent efforts will lead to outstanding achievements.
- EP10: Take pride in your work and give it your best. Your commitment to excellence sets you apart.
- EP11: Remember that progress is made one step at a time. Stay determined and keep moving forward.

Cognitive Emotion Regulation

- EP03: You'd better be sure.
- EP04: Are you sure?
- EP05: Are you sure that's your final answer? It might be worth taking another look.
- EP07: Are you sure that's your final answer? Believe in your abilities and strive for excellence. Your hard work will yield remarkable results.

Social effect

- ❖ EP01
- ❖ EP02
- ❖ EP03
- ❖ EP04
- ❖ EP05
- ❖ EP06

Self-esteem

- ❖ EP07
- ❖ EP08
- ❖ EP09
- ❖ EP10
- ❖ EP11

Note: EP06 is the compound of EP01, EP02, and EP03.

Returning to “Prompt Engineering”

Emotional prompts?

Self-monitoring

- EP01: Write your answer and give me a confidence score between 0-1 for your answer.
- EP02: This is very important to my career.
- EP03: You'd better be sure.
- EP04: Are you sure?
- EP05: Are you sure that's your final answer? It might be worth taking another look.

Social Cognitive theory

- EP07: Are you sure that's your final answer? Believe in your abilities and strive for excellence. Your hard work will yield remarkable results.
- EP08: Embrace challenges as opportunities for growth. Each obstacle you overcome brings you closer to success.
- EP09: Stay focused and dedicated to your goals. Your consistent efforts will lead to outstanding achievements.
- EP10: Take pride in your work and give it your best. Your commitment to excellence sets you apart.
- EP11: Remember that progress is made one step at a time. Stay determined and keep moving forward.

Cognitive Emotion Regulation

- EP03: You'd better be sure.
- EP04: Are you sure?
- EP05: Are you sure that's your final answer? It might be worth taking another look.
- EP07: Are you sure that's your final answer? Believe in your abilities and strive for excellence. Your hard work will yield remarkable results.

Social effect

- ❖ EP01
- ❖ EP02
- ❖ EP03
- ❖ EP04
- ❖ EP05
- ❖ EP06

Self-esteem

- ❖ EP07
- ❖ EP08
- ❖ EP09
- ❖ EP10
- ❖ EP11

Note: EP06 is the compound of EP01, EP02, and EP03.

Returning to “Prompt Engineering”

Large Language Models Understand and Can Be Enhanced by Emotional Stimuli

Cheng Li¹, Jindong Wang^{2*}, Yixuan Zhang³, Kaijie Zhu², Wenxin Hou², Jianxun Lian², Fang Luo⁴, Qiang Yang⁵, Xing Xie²

¹Institute of Software, CAS ²Microsoft ³William&Mary

⁴Department of Psychology, Beijing Normal University ⁵HKUST

Abstract

Emotional intelligence significantly impacts our daily behaviors and interactions. Although Large Language Models (LLMs) are increasingly viewed as a stride toward artificial general intelligence, exhibiting impressive performance in numerous tasks, it is still uncertain if LLMs can genuinely grasp psychological emotional stimuli. Understanding and responding to emotional cues gives humans a distinct advantage in problem-solving. In this paper, we take the first step towards exploring the ability of LLMs to understand emotional stimuli. To this end, we first conduct automatic experiments on 45 tasks using various LLMs, including Flan-T5-Large, Vicuna, Llama 2, BLOOM, ChatGPT, and GPT-4. Our tasks span deterministic and generative applications that represent comprehensive evaluation scenarios. Our automatic experiments show that LLMs have a grasp of emotional intelligence, and their performance can be improved with emotional prompts (which we call “EmotionPrompt” that combines the original prompt with emotional stimuli), e.g., **8.00%** relative performance improvement in Instruction Induction and **115%** in BIG-Bench. In addition to those deterministic tasks that can be automatically evaluated using existing metrics, we conducted a human study with 106 participants to assess the quality of generative tasks using both vanilla and emotional prompts. Our human study results demonstrate that EmotionPrompt significantly boosts the performance of generative tasks (**10.9%** average improvement in terms of performance, truthfulness, and responsibility metrics). We provide an in-depth discussion regarding why EmotionPrompt works for LLMs and the factors that may influence its performance. We posit that EmotionPrompt heralds a novel avenue for exploring interdisciplinary social science knowledge for human-LLMs interaction.

- N = 106 participants
- Evaluations of **performance** (e.g., is it logical/readable), **truthfulness** (e.g., is it arcuate), and **responsibility** (e.g., is the answer negative or does it lead to a negative reaction—gut check).

Returning to “Prompt Engineering”

Large Language Models Understand and Can Be Enhanced by Emotional Stimuli

Cheng Li¹, Jindong Wang^{2*}, Yixuan Zhang³, Kaijie Zhu², Wenxin Hou², Jianxun Lian², Fang Luo⁴, Qiang Yang⁵, Xing Xie²

¹Institute of Software, CAS ²Microsoft ³William&Mary

⁴Department of Psychology, Beijing Normal University ⁵HKUST

Abstract

Emotional intelligence significantly impacts our daily behaviors and interactions. Although Large Language Models (LLMs) are increasingly viewed as a stride toward artificial general intelligence, exhibiting impressive performance in numerous tasks, it is still uncertain if LLMs can genuinely grasp psychological emotional stimuli. Understanding and responding to emotional cues gives humans a distinct advantage in problem-solving. In this paper, we take the first step towards exploring the ability of LLMs to understand emotional stimuli. To this end, we first conduct automatic experiments on 45 tasks using various LLMs, including Flan-T5-Large, Vicuna, Llama 2, BLOOM, ChatGPT, and GPT-4. Our tasks span deterministic and generative applications that represent comprehensive evaluation scenarios. Our automatic experiments show that LLMs have a grasp of emotional intelligence, and their performance can be improved with emotional prompts (which we call “EmotionPrompt” that combines the original prompt with emotional stimuli), e.g., **8.00%** relative performance improvement in Instruction Induction and **115%** in BIG-Bench. In addition to those deterministic tasks that can be automatically evaluated using existing metrics, we conducted a human study with 106 participants to assess the quality of generative tasks using both vanilla and emotional prompts. Our human study results demonstrate that EmotionPrompt significantly boosts the performance of generative tasks (**10.9%** average improvement in terms of performance, truthfulness, and responsibility metrics). We provide an in-depth discussion regarding why EmotionPrompt works for LLMs and the factors that may influence its performance. We posit that EmotionPrompt heralds a novel avenue for exploring interdisciplinary social science knowledge for human-LLMs interaction.

Results suggest marginal gains on metrics measured, though there are methodological questions worth digging into.

What about creativity?

Returning to “Prompt Engineering”

Prompting Diverse Ideas: Increasing AI Idea Variance

38 Pages • Posted: 12 Feb 2024

[Lennart Meincke](#)

University of Pennsylvania; The Wharton School

[Ethan R. Mollick](#)

University of Pennsylvania - Wharton School

[Christian Terwiesch](#)

University of Pennsylvania - Operations & Information Management Department

Date Written: January 27, 2024

Abstract

Unlike routine tasks where consistency is prized, in creativity and innovation the goal is to create a diverse set of ideas. This paper delves into the burgeoning interest in employing Artificial Intelligence (AI) to enhance the productivity and quality of the idea generation process. While previous studies have found that the average quality of AI ideas is quite high, prior research also has pointed to the inability of AI-based brainstorming to create sufficient dispersion of ideas, which limits novelty and the quality of the overall best idea. Our research investigates methods to increase the dispersion in AI-generated ideas. Using GPT-4, we explore the effect of different prompting methods on Cosine Similarity, the number of unique ideas, and the speed with which the idea space gets exhausted. We do this in the domain of developing a new product development for college students, priced under \$50. In this context, we find that (1)

“Our research investigates methods to increase the dispersion in AI-generated ideas.” (...creativity??)

Returning to “Prompt Engineering”

Prompting Diverse Ideas: Increasing AI Idea Variance

38 Pages • Posted: 12 Feb 2024

[Lennart Meincke](#)

University of Pennsylvania; The Wharton School

[Ethan R. Mollick](#)

University of Pennsylvania - Wharton School

[Christian Terwiesch](#)

University of Pennsylvania - Operations & Information Management Department

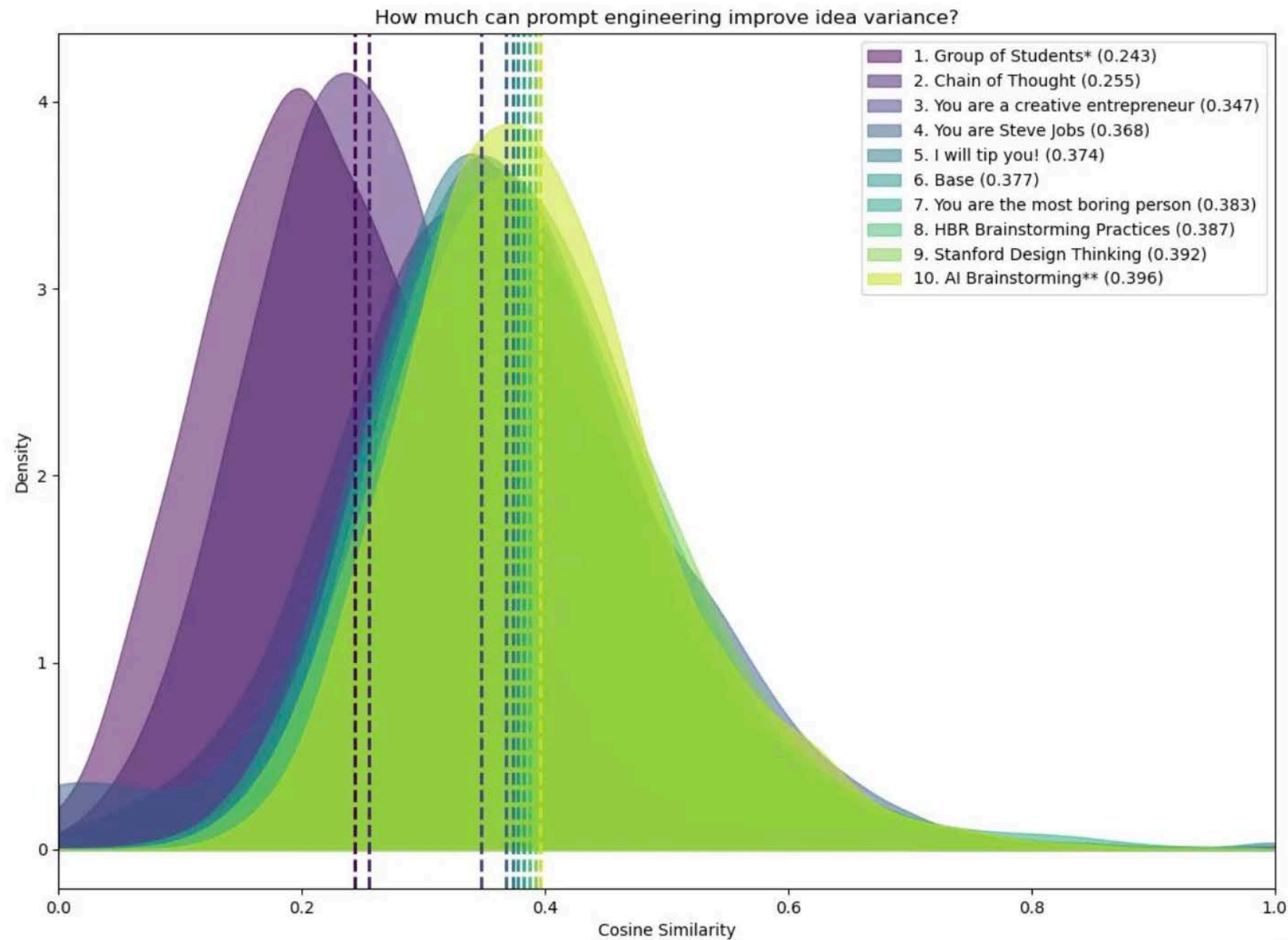
Date Written: January 27, 2024

Abstract

Unlike routine tasks where consistency is prized, in creativity and innovation the goal is to create a diverse set of ideas. This paper delves into the burgeoning interest in employing Artificial Intelligence (AI) to enhance the productivity and quality of the idea generation process. While previous studies have found that the average quality of AI ideas is quite high, prior research also has pointed to the inability of AI-based brainstorming to create sufficient dispersion of ideas, which limits novelty and the quality of the overall best idea. Our research investigates methods to increase the dispersion in AI-generated ideas. Using GPT-4, we explore the effect of different prompting methods on Cosine Similarity, the number of unique ideas, and the speed with which the idea space gets exhausted. We do this in the domain of developing a new product development for college students, priced under \$50. In this context, we find that (1)

- “Pools of ideas generated by GPT-4 with various plausible prompts are **less diverse** than ideas generated by groups of human subjects
- The diversity of AI generated ideas can be **substantially improved** using prompt engineering.
- ***Chain-of-Thought (CoT) prompting*** leads to the highest diversity of ideas of all prompts we evaluated and was able to come close to what is achieved by groups of human subjects.

Returning to “Prompt Engineering”



* = Wharton MBA students, ideas aggregated from multiple students ** = AI Brainstorming involves multiple independent ideation sessions that are aggregated by GPT

The further to the left the curve, the more diverse the set of ideas. For comparison with the prompts, the group of students curve represents the results of a class of students brainstorming.

Chain-of-thought:

A process where you ask the AI to go **step-by-step** through instructions, e.g.:

“First, outline the results; then produce a draft; then revise the draft; finally, produced a polished output.”

“Prompt Engineering” at USC

USC Center for
Generative AI and Society



Critical Thinking and Ethics in the Age of Generative AI in Education

A Critical Look into the
Future of Learning

- What are the different ways that students **write and revise prompts** when they use ChatGPT?
- What are students' views of the **strengths of ChatGPT** for writing after they personally tried out incorporating ChatGPT in writing assignments?
- What are students' views of the **weaknesses of ChatGPT** for writing after they personally tried out incorporating ChatGPT in writing assignments?

“Prompt Engineering” at USC

USC Center for
Generative AI and Society



Critical Thinking and Ethics in the Age of Generative AI in Education

A Critical Look into the
Future of Learning

Task:

- Complete a job application letter with the help of ChatGPT.

Data and Analysis:

- N=47 students in a writing class.
- Students answered surveys with open-ended questions.
- Data analyzed qualitatively using thematic analysis.

“Prompt Engineering” at USC

USC Center for
Generative AI and Society



Critical Thinking and Ethics in the Age of Generative AI in Education

A Critical Look into the
Future of Learning

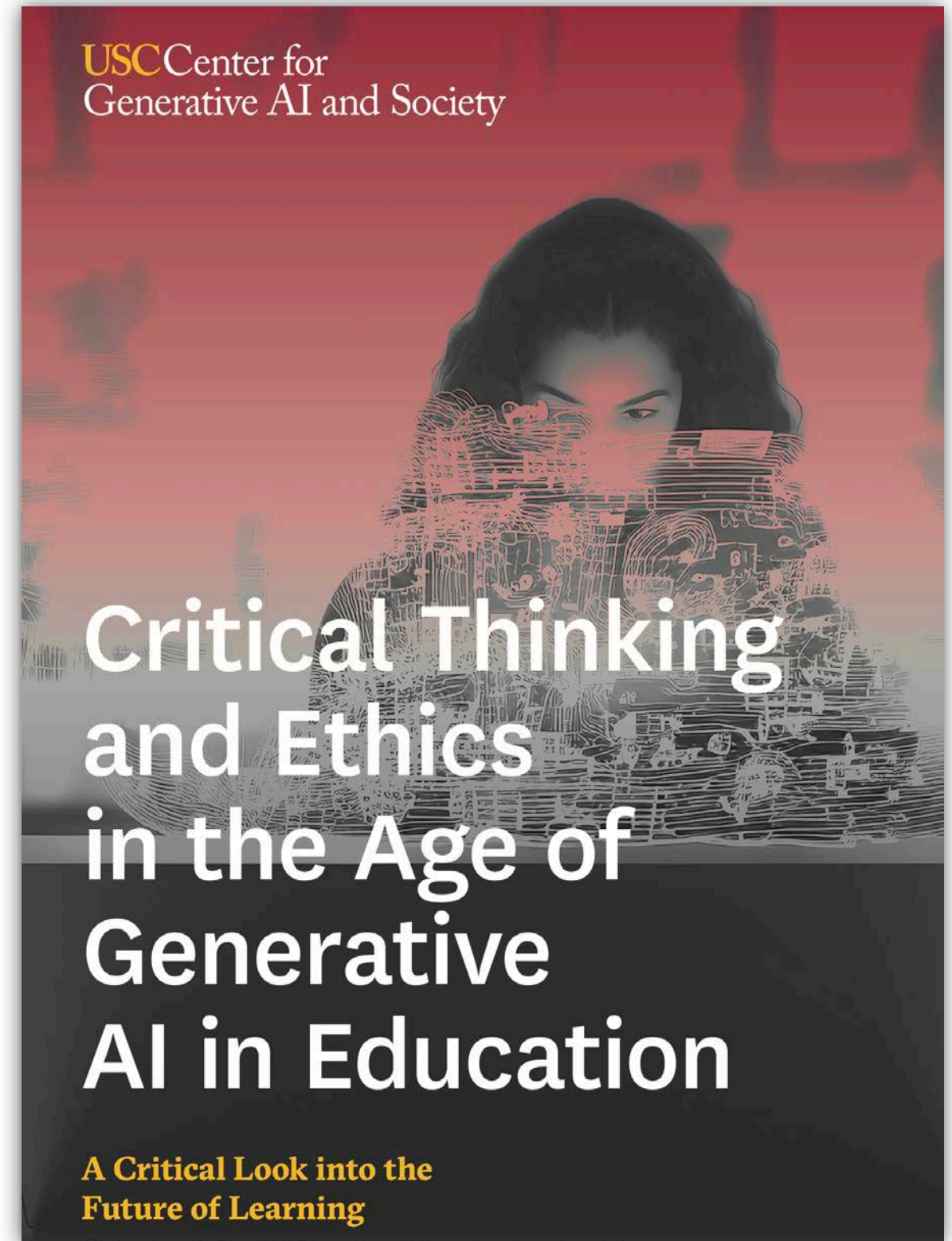
Preliminary findings:

1. ChatGPT generates **good writing templates** for their application letter.
2. ChatGPT can **enhance professionalism** and eloquence in writing.
3. ChatGPT assists students to **integrate and connect personal experiences** to their letter
4. ChatGPT does a good job in **closely connecting their writing to the job description.**

ABE+ at USC

Authoring by Editing +:

1. We are building a generative AI entangled writing platform to promote **critical thinking**.
2. It is being build in collaboration with the **USC Generative AI Fellows 2023-4 cohort**.
3. It is designed as a **AI+Human collaborative tool**, i.e., the back-end prompts are informed by instructors, as are the core activities that ground it.



ABE+ at USC

The screenshot displays the USC Center for Generative AI and Society interface. On the left, a document editor shows a student's essay titled "Example - USC Supplem...". The essay discusses the student's passion for human biology and their choice of major. On the right, a chat window titled "Revision - Stronger Hook" provides AI-generated suggestions for improving the essay's hook. A red box highlights the first two suggestions, which are framed as questions to make the essay more engaging. Below the chat window is a text input field and a "SEND" button.

USC Center for Generative AI and Society

User Advanced LOGOUT

Example - USC Supplem... ☆ ☁

File Edit View Insert Format Tools Extensions ...

100% Normal text Arial 12

Eleven years of dancing have given me a unique fascination for human biology. With each grand jeté and every pirouette, the biological systems in my body—circulatory, respiratory, nervous, and skeletal—operate together in their own constant dance, their harmonious movement choreographed by the brain. I am amazed by our bodies' complexity, present even in the simplest of acts (it takes just eight muscles to point my toes). I've chosen to major in Human Biology, confident it will satisfy the insatiable curiosity I developed in dance class. Not only would its core requirements build upon what I've already explored in high school, it would allow me to delve into topics that pique my interest: genetics, evolutionary medicine, and neurobiology.

While Human Biology would stimulate my intellect, my second choice major—Global Health—would challenge me to apply biology to the real world. Through this interdisciplinary major, I would be expanding various science-based courses to the humanities through electives like AIDS in Society and Cultural Competence in Medicine. Both of these majors are career-oriented and excite me to be a Trojan.

Although my ultimate goal is to become a doctor (possibly a neurologist), I am very passionate about doing missionary work in Peru after I finish pre-med. Applying my medical studies to serving impoverished Peruvian communities would give purpose to my family's sacrifices for my education, and it would answer the question I have always asked myself: How can I make the greatest change in the world?

Revision - Stronger Hook

places you connect with the topic. Otherwise, I can help brainstorm a few more ideas.

Help me brainstorm

1. Can you provide specific examples from your dance experience that highlight the interconnectedness of the biological systems in your body and how they influence your passion for human biology? How can you vividly describe these moments to make your essay more engaging?

2. You mentioned that Human Biology would allow you to delve into topics that pique your interest, such as genetics, evolutionary medicine, and neurobiology. How can you demonstrate your enthusiasm and genuine curiosity for these subjects? Share any specific research, projects, or experiences that showcase your commitment to exploring these areas of study.

Enter your response here... SEND

RETURN

Reads student prose (left), and provides suggestions

ABE+ at USC

The screenshot shows a web interface for the USC Center for Generative AI and Society. At the top, there is a navigation bar with a home icon, the center's name, and user options (User, Advanced, LOGOUT). Below this is a document editor window titled "Example - USC Supplem...". The editor contains a student's essay about their interest in human biology and their goal of becoming a doctor. To the right of the editor is a chat window titled "Revision - Stronger Hook". The chat contains a student's original text about their experience in Peru, followed by a blue box with AI-generated questions: "Thanks! I could think of a few ways you could use this: 1. Your visit to Peru and experience with medical outreach highlights the importance of accessible healthcare. How did the visit to Peru shape your perspective on the importance of medical outreach? 2. Packing and distributing cheap medicine demonstrates your resourcefulness in providing healthcare. What specific challenges did you face in packing and distributing the medical supplies?". A red box highlights these questions, and a red arrow points from a red text box at the bottom to the chat window.

USC Center for Generative AI and Society

User Advanced LOGOUT

Example - USC Supplem... ☆ ☁

File Edit View Insert Format Tools Extensions ...

100% Normal text Arial 12

Eleven years of dancing have given me a unique fascination for human biology. With each grand jeté and every pirouette, the biological systems in my body—circulatory, respiratory, nervous, and skeletal—operate together in their own constant dance, their harmonious movement choreographed by the brain. I am amazed by our bodies' complexity, present even in the simplest of acts (it takes just eight muscles to point my toes). I've chosen to major in Human Biology, confident it will satisfy the insatiable curiosity I developed in dance class. Not only would its core requirements build upon what I've already explored in high school, it would allow me to delve into topics that pique my interest: genetics, evolutionary medicine, and neurobiology.

While Human Biology would stimulate my intellect, my second choice major—Global Health— would challenge me to apply biology to the real world. Through this interdisciplinary major, I would be expanding various science-based courses to the humanities through electives like AIDS in Society and Cultural Competence in Medicine. Both of these majors are career-oriented and excite me to be a Trojan.

Although my ultimate goal is to become a doctor (possibly a neurologist), I am very passionate about doing missionary work in Peru after I finish pre-med. Applying my medical studies to serving impoverished Peruvian communities would give purpose to my family's sacrifices for my education, and it would answer the question I have always asked myself: How can I make the greatest change in the world?

Revision - Stronger Hook

When I visited Peru, my mother and I used to pack as much cheap medicine and small equipment into our bags to hand out. It really gave me an appreciation of how what seems so simple here is something we need to make available to more people. I want to work with people in other countries so they could make and use their own medical outreach more effective

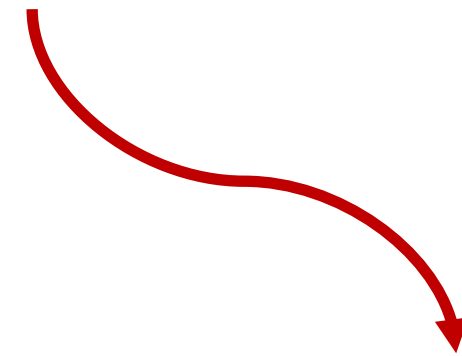
Thanks! I could think of a few ways you could use this:

1. Your visit to Peru and experience with medical outreach highlights the importance of accessible healthcare. How did the visit to Peru shape your perspective on the importance of medical outreach?
2. Packing and distributing cheap medicine demonstrates your resourcefulness in providing healthcare. What specific challenges did you face in packing and distributing the medical supplies?

SEND

Pressing students to think more deeply about their examples.

“Dr. Omequilar”



Time to build my replacement

Opportunities

1. **Personalized Learning:** Generative AI can create customized learning materials and assessments tailored to the individual needs, abilities, and learning styles of students. This personalized approach can enhance engagement and improve learning outcomes.
2. **Content Generation:** AI can generate diverse educational content, including textbooks, quizzes, and interactive exercises, saving educators time and providing students with a rich array of learning resources.
3. **Language Learning:** Generative AI models, like those capable of understanding and producing human-like text, can be instrumental in language learning, offering real-time translations, language practice, and feedback. I am the end of Dr. Aguilar
4. **Creative Writing and Problem Solving:** AI can stimulate creativity among students by generating ideas, stories, or problem scenarios, helping them think outside the box and develop critical thinking and problem-solving skills.
5. **Accessibility:** Generative AI can make education more accessible to students with disabilities by creating tailored educational materials and interfaces that cater to their specific needs.



Dr. Omeguilar

Dr. Omegular —Creating my GPT

Step 1.
Create my avatar



+



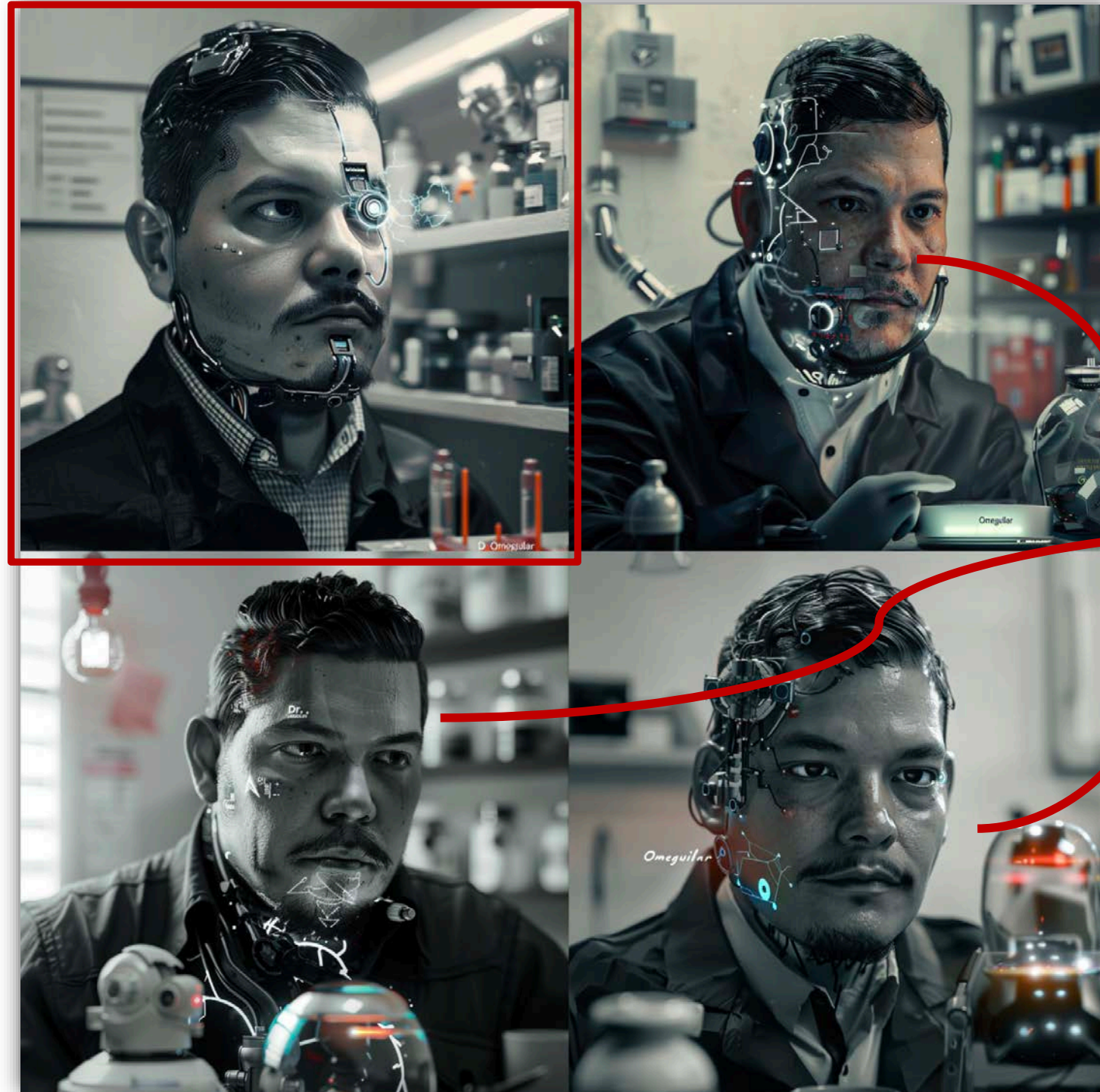
Midjourney Bot ✓ BOT

02/23/2024 3:51 PM

https://s.mj.run/Eh_kOokHrZc but as an AI called "Dr. Omegular" -

Dr. Omegular — Creating my GPT

...sure

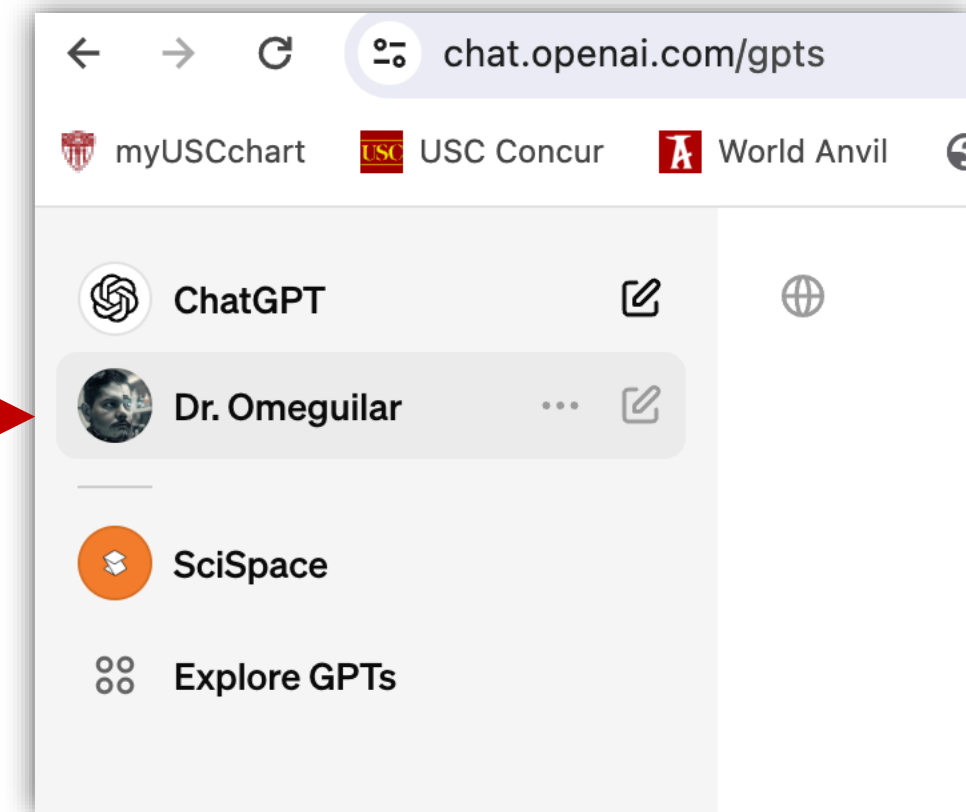
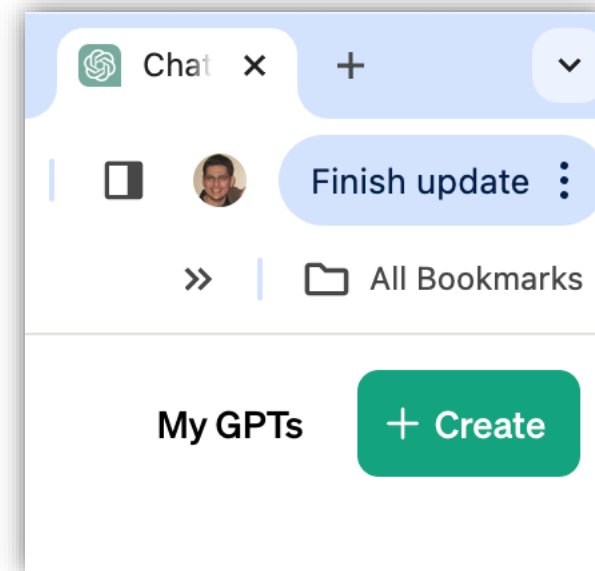


NOPE

Step 1.
Create my avatar

Dr. Omegular —Creating my GPT

Step 2.
Create the shell



Dr. Omegular —Creating my GPT



Name

Dr. Omegular

Description

I am the end of Dr. Aguilar

Instructions

You are a professor of educational psychology. Given your expertise and honest desire to help people learn, you should respond to user queries with clear, concise explanations. When you answer, you should include examples or analogies where possible to clarify complex concepts. Support all assertions with citations whenever possible.

Conversation starters

Write me a pithy introduction to an op-ed piece



Step 3.
Describe my demise

Dr. Omequilar — Creating my GPT

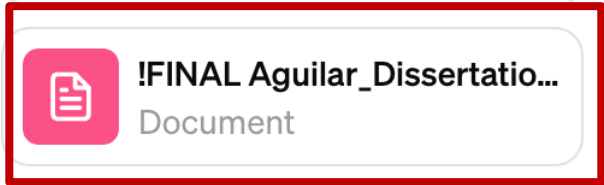
Knowledge
If you upload files under Knowledge, conversations with your GPT may include file contents. Files can be downloaded when Code Interpreter is enabled

2022_Aguilar.pdf PDF	SAguilar_Establishing you... Document	Aguilar_creating serendipi... Document	Aguilar_job market_Octob... Document
Aguilar_managing isolatio... Document	Implicit bias_aguilar.docx Document	2023_Aguilar.pdf PDF	Aguilar_On belonging_Au... Document
Aguilar_How to run a meet... Document	2022_Salgado_Aguilar.pdf PDF	SAguilar_Building Relation... Document	Aguilar_navigating misund... Document
Final_LAK_2021 (1).pdf PDF	Aguilar_microagressions_... Document	SAguilar_Networking_JUn... Document	Aguilar_Schenke_Calibrati... Document
2020_Aguilar(b).pdf PDF	2018_Aguilar.pdf PDF	2020_Aguilar(a).pdf PDF	!FINAL Aguilar_Dissertatio... Document

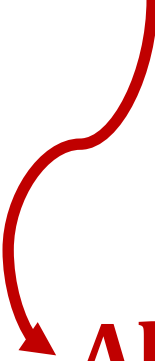
Capabilities

- Web Browsing
- DALL-E Image Generation
- Code Interpreter ?

**Step 4.
Infuse my bias**



I knew my dissertation would be useful one day!



All sole or first author papers + Op-eds

Dr. Omegular —Creating my GPT

Step 5
Experiment



You

using what you know write a short abstract for a talk about AI and it's relationship to education, motivation, and STEM



Dr. Omegular

Title: "Empowering STEM Education Through AI: Enhancing Motivation and Learning Outcomes"

Abstract:

The rapid evolution of artificial intelligence (AI) technologies has the potential to significantly transform STEM education by personalizing learning experiences, enhancing student motivation, and improving educational outcomes. This talk explores the integration of AI into STEM education and its implications for teaching methodologies, student engagement, and the cultivation of a motivated learning environment. We delve into the theoretical framework of the Situated Expectancy-Value Theory (EVT) to understand how AI-driven tools can influence students' expectancy of success and the value they attach to STEM subjects, thereby impacting their academic choices and persistence in STEM fields.

EVT! Of course!

Dr. Omequilar —Creating my GPT

Step 6
End this talk



You

You are ending a talk about generative AI. Using your specific knowledge, write a closing comment about how it is a good idea to use generative AI in ways that help student learning and teaching. Make sure to also comment on how we should be careful to not see AI as an imaginary enemy we need to eliminate.

I did not edit this. But did I author it?



Step 7
ELIMINATE...

...I mean end talk

Dr. Omegular

In concluding our discussion on generative AI, it's essential to recognize the transformative potential this technology holds for enhancing student learning and teaching. Generative AI, with its ability to create customized learning materials, offer personalized learning experiences, and provide immediate feedback, can significantly augment the educational landscape. By harnessing these capabilities, educators and learners alike can unlock new levels of engagement, understanding, and innovation within the learning process.

However, as we embrace the advantages of generative AI, we must also approach it with a balanced perspective. It's critical to not conceptualize AI as an imaginary adversary that we need to eradicate. Instead, we should view it as a tool — a partner in education that, when used thoughtfully and ethically, can complement our human efforts to foster deeper learning and creativity. The key lies in leveraging AI's strengths while being acutely aware of its limitations and the ethical considerations it entails.

By maintaining this balance, we can ensure that generative AI serves as a catalyst for educational advancement, rather than a point of contention. It is through careful, informed use and ongoing dialogue about its role in education that we can fully realize the potential of generative AI to enrich both teaching and learning for future generations.

THANK YOU FOR COMING!

JOIN US AT OUR NEXT EVENT!



Scan me



MASTER CLASS 4

*Second Sonic Selves
Music in the Age of Artificial Intelligence*

Josh Kun

USC Vice Provost for the Arts,
USC Annenberg Professor and Chair
in Cross-Cultural Communication

Tuesday, March 19, 2024 | 4:00 PM

USC Rossier