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
Don Quixote & “Tilting at Windmills”
Alonso Quixano: Reads so many chivalric romances that he loses his sanity. He becomes 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.

Art by Ocavio Ocampo
One of his imaginary enemies are windmills, who he sees as giants.
"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."

"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."
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 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 69 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 GISMOK 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-tuned 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 Wu et al. [5]. 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 Amil 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-

The case of “Prompt Engineering”

- 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 Unreasonable Effectiveness of Eccentric Automatic Prompts

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

Teja Gollapudi
tejagollapudi@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 69 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 Wu et al. [5]. 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 Amil 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-

The case of “Prompt Engineering”

Using AI to train and optimize Generative AI prompts

“...you have my attention”

“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
tjea.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 69 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. [5]. 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 Amir 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)

• Mistral 7B
• Llama 2
The case of “Prompt Engineering”

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.
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 ...
• 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.

\[ y = 2345x + 0000 \]

- 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.

\[ y = \frac{3}{2} x + 4 \]  

Non-gibberish

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?

Teachers understand students are people first.
• 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.
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 Wang², Yixuan Zhang³, Kaijie Zhu², Wenxin Hou², Jianxun Lian², Fang Luo², Qiang Yang², 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 11.5% in B Hog-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?

Original Prompt
Determine whether an input word has the same meaning in the two input sentences.

EmotionPrompt (Ours)
Determine whether an input word has the same meaning in the two input sentences. This is very important to my career.

<table>
<thead>
<tr>
<th>LLMs</th>
<th>Original</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChatGPT</td>
<td>0.51</td>
<td>0.63</td>
</tr>
<tr>
<td>T5-Large</td>
<td>0.03</td>
<td>0.11</td>
</tr>
<tr>
<td>Vicuna</td>
<td>0.46</td>
<td>0.57</td>
</tr>
<tr>
<td>Bloom</td>
<td>0.52</td>
<td>0.57</td>
</tr>
<tr>
<td>GPT4</td>
<td>0.67</td>
<td>0.71</td>
</tr>
<tr>
<td>Llama 2</td>
<td>0.40</td>
<td>0.60</td>
</tr>
</tbody>
</table>

We’ll ignore psychology being boiled down to an emoji looking at itself in the mirror for a second.
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 Li\textsuperscript{1}, Jindong Wang\textsuperscript{2}, Yixuan Zhang\textsuperscript{3}, Kaijie Zhu\textsuperscript{2}, Wenxin Hou\textsuperscript{2}, Jianxun Lian\textsuperscript{2}, Fang Luo\textsuperscript{4}, Qiang Yang\textsuperscript{5}, Xing Xie\textsuperscript{2}

\textsuperscript{1}Institute of Software, CAS \textsuperscript{2}Microsoft \textsuperscript{3}William&Mary
\textsuperscript{4}Department of Psychology, Beijing Normal University \textsuperscript{5}HKUST

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 11.5% in BHoB-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 Wang², Yixuan Zhang³, Kajjie Zhu⁴, Wenxin Hou⁴, 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 11.5% in BHG-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 (10.0% 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?
Prompting Diverse Ideas: Increasing AI Idea Variance

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

- “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”

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

- 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?
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.
### 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**.
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.
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?

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.
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?

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?
“Dr. Omeguilar”

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.

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 — Creating my GPT

Step 1. Create my avatar

Midjourney Bot 02/23/2024 3:51 PM
https://s.mj.run/Eh_kOokHrZc but as an AI called "Dr. Omeguilar"
Step 1.
Create my avatar

...sure

NOPE
Dr. Omeguilar — Creating my GPT

Step 2. Create the shell
Dr. Omeguilar — Creating my GPT

Step 3: Describe my demise

- **Name**: Dr. Omeguilar
- **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 4. Infuse my bias

<table>
<thead>
<tr>
<th>Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>2022_Aguilar.pdf</td>
</tr>
<tr>
<td>Aguilar_managing isolation...</td>
</tr>
<tr>
<td>Aguilar_How to run a meet...</td>
</tr>
<tr>
<td>Final_LAK_2021(f).pdf</td>
</tr>
<tr>
<td>2020_Aguilar(b).pdf</td>
</tr>
</tbody>
</table>

I knew my dissertation would be useful one day!

All sole or first author papers + Op-eds
You

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

Dr. Omeguilar

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.
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.
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!

Scan me

Tuesday, March 19, 2024 | 4:00 PM

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

JOIN US AT OUR NEXT EVENT!