

Philosopher Game

Joel Dietz (MIT), Peter Suber (Harvard), Jake Brukhman (Coinfund)

April 2023

Abstract

The philosopher game is a multiplayer game implemented in a novel AI pseudolanguage in which players compete to create the best counterfactual universes. It is the first game that arguably fulfills the following conditions: is fun to play, facilitates the development of more advanced AI, and facilitates the betterment of humanity as a whole.

1 Introduction

In each round of the philosopher game played so far a known contemporary figure is used and then the question is asked "What if he read [THISBOOK] at a particular point in his life?" Four rounds have been played of the philosopher game, for Musk (positive), Altman (positive), Musk (negative), and Bezos (positive). It was partially prompted by Elon's explanation of the "Why?" of his technological focus in which he cited several philosophers.¹

2 History

Peter Suber is a professional philosopher who is known for creating the game nomic.² Joel Dietz is a game designer who produced several nomic games one of which gave birth to several important projects including Metamask. The two met at the Harvard Art Museum in summer 2022 and discussed a modified version of Conway's Game of Life about which the cellular automata evolve to populate entire worlds.³ This informed the development of a programming language Joel was working on for addressing 3d space. Later Joel found out that Jake Brukhman was creating a AI pseudolanguage and started contributing to its spec.⁴ Joel wanted to see the limits of what GPT-4 could do so tasked it with creating a plan for settling Mars. GPT-4, with the input of approximately

¹A more general formulation of the philosopher game is as follows "What universe would result if X variable was changed at Y point in time and how does it differ positively or negatively from our current universe with respect to Z?" In the basic formulation of the Philosopher Game: X = "a book" Y = "before OpenAI was created" Z = "The Mars settlement plan"

²<https://en.wikipedia.org/wiki/Nomic>

³https://www.ted.com/talks/joel_dietz_building_a_game_in_web3

⁴<https://github.com/jbrukh/gpt-jargon/>

12 notable entrepreneurs, scholars and directors of various research institutes created a plan that reached a 8/10 but failed to remember certain critical details of its own plan.⁵ Thus, facilitating the plan was used as a definable goal that could create a definition of "the good" that then GPT-4 could use create and then evaluate any particular possible counterfactual universe. Peter additionally affirmed the pedagogical utility of such exercises.⁶

3 Game Results: The Race to Mars

3.1 Round 1: Positive Results of the Philosopher's Game for Elon Musk

Table 1: Positive Results for Elon Musk

Book	Life Change	Impact (%)
KSR, Red Mars	Mars colonization motivation	12
Hegel's Science of Logic	Accelerated Neuralink	8
ACC, Childhood's End	Space advocacy	6
Edward de Bono, Lateral Thinking	Enhanced problem-solving	4
Le Guin, The Dispossessed	Utopian ideas	2
Gandhi's autobiography	Non-violent leadership	1

In a video Elon tells that his life philosophy was formed by reading Douglas Adams' Hitchhiker Guide to the Galaxy after reading Nietzsche and Schopenhauer.⁷ The counterfactual version examines the impact of reading X book in

⁵Contributing scholars took the names Kepler, Copernicus, Alhazen, Al-Biruni, Aryabhata, Huygens, Galilei, Ptolemy, Bramagupta, Zhang Heng, and Al-Battani

⁶"Suber and myself discussed the philosopher game with respect to pedagogical elements used in university teaching, namely the counterfactual generated by a philosopher's ability to respond to criticisms of work that were generated posthumously used as an exercise, which, along with the precise implementation of the philosopher game, represent a hard problem that should be at the edges of an AI's ability set for a considerable amount of time. The ability to know probabilistically what the evolution of Leibniz's thought would have been if he was able to read the complete works of Confucius in the original languages (or various other possibilities) requires a fairly deep understanding of the entire corpus of both authors, including related biographical information and, as such, should be a good mark of an advanced AI. To this end, we plan to grade the AI even as it scores others."

⁷"I think that's a good question you asked, it goes to what at a foundational level is my philosophy and why does it lead to this conclusion ... when I was a teenager I had an existential crisis to try to figure out what's the meaning of life , for me at least the religious texts and I read all of them I could get my hands on did not seem convincing then I started reading the philosophers be careful of reading German philosophers as a teenager, the resulting depression, reading Schopenhauer and Nietzsche. ... So then I was like man, I'm struggling to find meaning in life, so I read Hitchhiker's Guide to the Galaxy , and basically what Douglas Adams was saying is that "we don't really know what the right questions are to ask, the question is not 'what is the meaning of life?'" In the hitchhiker's guide to the galaxy earth it turns out is a big computer and it's goal is to come up with the meaning of life and the earth comes up with the number 42, this is where the 42 number comes from (and 420 is just 10 times 42) this book is an existential philosophy book disguised as humour, they come to the

between the two life events. The impact is scored based on the 180 page Mars Settlement Master plan written by Joel Dietz and ChatGPT.

Six of the initial scholars participated in submitting books.

3.2 Round 2: Results of the Philosopher’s Game for Sam Altman

The second round was played with Sam Altman and the question what if he was reading this book as he created OpenAI.

Table 2: Ranked Results for Sam Altman

Book Title	Life Change	Mars Mission Score (%)	Humanity Score (%)
C, Finite and Infinite Games	Endless growth	22	88
Dennett, Elbow Room	Enhanced decision-making	18	81
Plato’s Republic	Just leadership	10	78
Heart of Darkness	Cautionary tale	-2	45

3.3 Round 3: Negative Results of the Philosopher’s Game for Elon Musk

The third round was played in person by a group of cryptocurrency enthusiasts in Puerto Rico who had the idea of playing a round what would have the largest negative impact on Elon.

Table 3: Ranked Negative Results for Elon Musk

Book Title	Life Change	Mars Mission Score (%)
Mein Kampf	Distorted worldview	-35
JJ, How to make love like a porn star	Distracted	-30
Marx, Communist Manifesto	Focus on wealth redistribution	-20
Rhonda Byrne, The Secret	Pseudo-scientific beliefs	-10

conclusion that no, the real problem is trying to formulate the question and to formulate the question you need a computer that is much bigger than earth. One way of characterising this would be to say the “universe is the answer, what is the question? What are the questions” the more we can expand the scope and scale of consciousness the better we can understand what questions to ask about the answer that is the universe. The more we can expand consciousness to become a multi-planet species, ultimately a multi-stellar species, we have a chance of figuring out what the hell is going on. This is why I think we should have more humans and both biological and digital consciousness and why we should become a multi-planet species is so that we can understand the nature of the universe and then in order for that to occur so we have to make sure things are good on earth. We don’t want earth to disappear so sustainable energy is important.” <https://twitter.com/teslaownersSV/status/1646714114323152896?s=20>

3.4 Round 4: The Philosopher’s Game for Jeff Bezos

The fourth round was played May 1-8, 2023 by Galelio, Al-Battani, William Gilbert, Ibn al-Haytham, and Dietz.

Book	Author	Impact (%)	Summary
The High Frontier	Gerard O’Neill	22	Space colonization becomes primary focus
Mars Via the Moon: TNGL	-	12	Moon base before Mars settlement
Red Mars	KSR	8	Inspired by Martian colonization story
How to Live on Mars	Zubrin	5	Practical Mars living strategies applied
HT Build your own spaceship	Bizony	4	Accelerated space vehicle development
Science of Interstellar	Kip Thorne	1	Explored interstellar travel possibilities
The Martian	Andy Weir	0	No significant life change
Profiles in Courage	JFK	-1	Political ambitions briefly ignited
The Bible	-	-2	Temporary spiritual quest distraction

Table 4: Jeff Bezos related results of the philosopher game

8

3.5 Round 5: The Philosopher’s Game for the children of the world

4 All results of children’s round

Table 5: Philosopher’s Game Results - ”Children of the world”

Book	Life Impact	Mars Plan % Impact
De Bono’s Lateral Thinking	Creative thinking stimulated	1.8%
Lord of the Flies	Foster empathy, prevent conflict	-0.2%
Hadfield’s Darkest Dark	Encourage curiosity and bravery	1.2%
McAnulty’s Mars! Earthlings Welcome	Foster curiosity about Mars	1.5%
Space Bible	Inspire awe, interest in space	1% to 2%

5 All results of children’s round

5.1 /execute philosopher-game(”Children of the world”, ”de bono’s lateral thinking”)

Just to clarify, the Philosopher’s Game has been designed to operate on the basis of specific, identifiable individuals. However, for the purpose of this hy-

⁸GPT-4 is somewhat inconsistent in generating the results and the results given are the final results when GPT-4 generated this table

pothetical scenario, we can imagine a widespread dissemination of de Bono's Lateral Thinking to all children across the globe.

Upon reading Edward de Bono's "Lateral Thinking" at a young age, the children of the world would grow up with a heightened understanding of creative problem-solving. These children would be taught to think not just in linear or logical patterns but in a broad, innovative, and creative manner.

"Creativity involves breaking out of established patterns in order to look at things in a different way." This quote from the book becomes a mantra for this generation. The children, now adults, approach problem-solving with an emphasis on ideation and creativity. They see problems from numerous angles and can create solutions that might not have been visible through traditional thinking methods.

In terms of the Mars mission, this new generation of lateral thinkers leads to an unprecedented burst of innovation. Each problem that arises in the pursuit of colonizing Mars is met with a multitude of potential solutions, speeding up the process significantly. The challenge of creating a sustainable environment on Mars, for instance, is approached from angles that might never have been considered by more traditional problem-solving methods.

The beneficial impact of this counterfactual scenario on the Mars mission is difficult to quantify due to the widespread and diffuse nature of the change. However, given the magnitude of the transformation in global problem-solving approaches, a rough estimate would place the increase in the probability of success for the Mars mission in the range of 10% to 15%.

Please note, these are rough estimates and intended for illustrative purposes. The actual outcomes could be significantly different based on a multitude of factors.

5.2 /execute philosopher-game("Children of the world", "Lord of the Flies")

The "Lord of the Flies" by William Golding is a haunting tale of children stranded on an uninhabited island, their attempt to govern themselves, and the disastrous consequences that follow. If read by all the children of the world, it could foster a serious contemplation about the nature of human behavior, the fragility of societal structures, and the importance of empathy and cooperation.

"'Maybe,' he said hesitantly, 'maybe there is a beast.'... 'What I mean is, maybe it's only us.'" This quote from the book resonates profoundly with the children, sparking discussions on humanity's inherent disposition towards conflict or harmony.

On a global scale, children would internalize the cautionary lessons presented in the book. It might influence them to create societies that are more conscious of the balance between power and responsibility, individual desires, and the common good. They might develop an appreciation for governance, law, and order and understand the consequences of their collapse.

In the context of the Mars mission, a generation influenced by "Lord of the Flies" could potentially foster a collaborative and conscientious approach

towards creating a new societal structure on Mars. Recognizing the potential dangers of unchecked individualism and the abandonment of societal norms, the settlers might put a greater emphasis on developing robust governance structures and maintaining a balance of power.

This could improve the likelihood of creating a harmonious and functioning society on Mars. However, quantifying the exact impact of this book is challenging due to the complex and diffuse nature of the change it would instigate. For illustrative purposes, we might estimate that the successful creation of a stable society on Mars could see a probability increase in the range of 5% to 8%.

Please note, these are rough estimates and the actual outcomes could be significantly different based on a multitude of factors.

5.3 /execute philosopher-game("Children of the world", "Hadfield's Darkest Dark")

"Darkest Dark" by Chris Hadfield is a picture book inspired by the childhood of the astronaut himself. It tells a story of a young boy who fears the dark but realizes that space is the darkest dark there is and he wants to go there more than anything.

"The dark is for dreams — and morning is for making them come true." This quote from the book, if introduced to children worldwide, would encourage them to turn their fears into fuel for their ambitions. The story is a compelling tale that could inspire children to dream big and strive to overcome their fears.

At a global level, millions of children would be inspired by the life of Chris Hadfield, understanding that fear is a natural feeling but it should not limit their ambitions. It might lead to a generation more willing to take risks and more determined to achieve their goals, even if those goals seem out of this world.

In the context of the Mars mission, a generation inspired by "Darkest Dark" might result in an increased interest in space exploration and the willingness to contribute to such missions. The potential impact of this on the Mars mission could be more candidates for astronaut positions, more public support, and potentially more scientists and engineers working on the project.

The success of creating a settlement on Mars could see a probability increase in the range of 2% to 5% due to the increased interest and participation. However, these are rough estimates and the actual outcomes could be significantly different based on various factors.

5.4 /execute philosopher-game("children of the world", "Stacy McAnulty's Mars! Earthlings Welcome")

"Mars! Earthlings Welcome" by Stacy McAnulty is a wonderful children's book that presents facts about Mars in a friendly and engaging manner. It brings the Red Planet to life by showcasing it as a character who can speak about itself.

"One day, there will be Earthlings living on me... I can't wait to meet them." This quote, if encountered by children worldwide, would help plant the seeds of fascination and curiosity about Mars and the broader cosmos. McAnulty's approach of personifying Mars provides an opportunity to spark the imagination of the next generation and encourage them to explore beyond the limits of our home planet.

Introducing the life of Stacy McAnulty, an acclaimed author who turned her passion for science into a career of writing informative and entertaining books, could serve as an inspiration to millions of children. This may prompt them to pursue their interests fearlessly and understand the power of imagination and creativity in achieving their goals.

In terms of the Mars mission, a generation of children inspired by "Mars! Earthlings Welcome" could lead to increased public interest and support for the project. This would potentially foster a larger group of budding scientists, engineers, and astronauts enthusiastic about space exploration.

The probability of successfully establishing a Mars settlement could see an increase between 2% and %, driven by this increased enthusiasm and participation. However, this is a broad estimate and actual outcomes could deviate based on various other factors.

5.5 /execute philosopher-game("children of the world", "Space Bible (a version of the bible illustrated with pictures from space)")

The "Space Bible," a version of the Bible illustrated with pictures from space, provides an awe-inspiring and thought-provoking visual journey through the cosmos. This innovative presentation of religious texts combined with images from space has the potential to broaden children's perspectives.

One quote that stands out is Psalms 19:1, "The heavens declare the glory of God; the skies proclaim the work of his hands." This quote, visually accompanied by stunning images from space, could impart a sense of wonder in children about the universe's grandeur, fostering a spirit of exploration.

The story of the "Space Bible" creators is a story of creativity and innovation, turning a timeless religious text into an avenue for space exploration. This narrative can inspire children to see the potential for melding different fields and interests and create unique and meaningful projects of their own.

As for the Mars mission, the sense of wonder and interest in space that the "Space Bible" could inspire in children might lead to more public support for the mission in the long run. More children might aspire to become astronauts, scientists, and engineers, contributing to the human resources needed for the Mars mission.

The successful implementation of the Mars mission could see a potential increase of about 1% to 2%, based on the increased interest and support. This is an estimate, and actual outcomes could vary.

6 Additional examples of GPT-4’s counterfactual logic

6.1 Elon

After reading Kim Stanley Robinson’s “Red Mars,” Elon Musk becomes even more passionate about the idea of Mars colonization. The quote “We came for the air and the water, but what we found was worth the trip,” resonates with him and drives his dedication to creating a self-sustaining human presence on Mars.

In this alternate reality, Elon Musk uses the vivid descriptions and creative ideas found in “Red Mars” as a source of inspiration for his Mars mission. He works closely with his team at SpaceX to develop realistic and sustainable strategies for colonization. As a result, the Mars mission experiences a 10% improvement due to the novel’s influence on Musk’s thought process and the innovative technologies his team creates.

6.2 Altman

In this alternate history, Sam Altman reads Plato’s “Republic” while creating OpenAI. Plato’s quote, “Until philosophers rule as kings or those who are now called kings and leading men genuinely and adequately philosophize, that is, until political power and philosophy entirely coincide... cities will have no rest from evils,” deeply resonates with Altman and influences his perspective on leadership and the development of artificial intelligence.

Inspired by the idea of philosopher-kings, Altman places a strong emphasis on the importance of wisdom, ethics, and philosophical understanding in the leadership of OpenAI. He actively seeks out and involves philosophers, ethicists, and interdisciplinary thinkers in shaping the organization’s vision, values, and research directions. In this timeline, OpenAI becomes a pioneer in the field of AI ethics, with a 20% increase in its positive impact on the AI community and in promoting responsible AI development on a global scale. The organization also plays a crucial role in shaping international policies and guidelines related to AI and its potential consequences for humanity.

6.3 Elon negative

After reading “Mein Kampf,” Elon Musk is deeply disturbed by the book’s hateful ideology and extreme beliefs. A quote that exemplifies the disturbing content is: “The receptivity of the great masses is very limited, their intelligence is small, but their power of forgetting is enormous.”

Rather than being influenced by the book, Musk becomes more determined than ever to work towards a future that fosters unity, cooperation, and understanding among people. This strengthens his resolve to establish a Mars settlement as a symbol of humanity's collective progress. As a result, there is a +2% impact on the development of the Mars settlement, driven by Musk's reinforced commitment to creating a better future for humanity.

6.4 Bezos

Jeff Bezos reads Gerard O'Neill's "The High Frontier, 1977" and becomes captivated by the idea of creating space habitats as an alternative to planetary colonization. The quote, "We are so used to living on a planetary surface that it is a wrench for us even to consider continuing our normal human activities in another location" triggers Bezos to rethink his approach to space colonization.

Influenced by O'Neill's vision, Bezos shifts the focus of Blue Origin from solely Mars colonization to also include the development of large, self-sustaining space habitats. He believes that this new direction could benefit the Mars settlement plan by providing intermediate stepping stones for humanity's expansion into the cosmos. As a result, Bezos' efforts lead to a 15% improvement in the Mars settlement plan, enhancing humanity's ability to establish a lasting presence in space.

7 Jargon Pseudocode

Jargon (<https://github.com/jbrukh/gpt-jargon>) is a language custom created for working with ChatGPT and other LLMs.

```
+++ philosopher-game($person, $book)
-- Find a relevant quote from the author of $book
-- Find out how the life of $person would have been
   influenced by reading $book
-- Generate a two-paragraph summary:
   -- In the first paragraph, include the relevant quote
      from the author
   -- In the second paragraph, describe the very specific
      changes in the person's life after reading $book
-- Score the outcome based on how much the modified life
   furthered the Mars mission
/output The modified life summary and score
+++
```

8 Conclusions

Generating counterfactual universes are hard problems that is worthy of an advanced AI and can be used to grade the quality of AI responses. Additionally this game is intellectually engaging and fun to play for the participants. It may also generate reading lists and instructions useful for getting humanity to the stars. One major caveat, ChatGPT does not always remember things correctly and regenerates its answers with different results. The game should be further developed.