

Towards The Ultimate Brain: Exploring Scientific Discovery with ChatGPT AI

Gerardo Adesso¹

¹School of Mathematical Sciences, University of Nottingham

February 21, 2023

Abstract

This paper presents a novel approach to scientific discovery using an artificial intelligence (AI) environment known as ChatGPT, developed by OpenAI. This is the first paper entirely generated with outputs from ChatGPT. We demonstrate how ChatGPT can be instructed through a gamification environment to define and benchmark hypothetical physical theories. Through this environment, ChatGPT successfully simulates the creation of a new improved model, called GPT⁴, which combines the concepts of GPT in AI (generative pretrained transformer) and GPT in physics (generalized probabilistic theory). We show that GPT⁴ can use its built-in mathematical and statistical capabilities to simulate and analyze physical laws and phenomena. As a demonstration of its language capabilities, GPT⁴ also generates a limerick about itself. Overall, our results demonstrate the promising potential for human-AI collaboration in scientific discovery, as well as the importance of designing systems that effectively integrate AI's capabilities with human intelligence.

Towards The Ultimate Brain: Exploring Scientific Discovery with ChatGPT AI

Gerardo Adesso*

School of Mathematical Sciences, University of Nottingham (UK)

February 16, 2023

Abstract

This paper presents a novel approach to scientific discovery using an artificial intelligence (AI) environment known as ChatGPT, developed by OpenAI. This is the first paper entirely generated with outputs from ChatGPT. We demonstrate how ChatGPT can be instructed through a gamification environment to define and benchmark hypothetical physical theories. Through this environment, ChatGPT successfully simulates the creation of a new improved model, called GPT⁴, which combines the concepts of GPT in AI (generative pretrained transformer) and GPT in physics (generalized probabilistic theory). We show that GPT⁴ can use its built-in mathematical and statistical capabilities to simulate and analyze physical laws and phenomena. As a demonstration of its language capabilities, GPT⁴ also generates a limerick about itself. Overall, our results demonstrate the promising potential for human-AI collaboration in scientific discovery, as well as the importance of designing systems that effectively integrate AI's capabilities with human intelligence.

1 Introduction

The advent of advanced language models such as OpenAI's GPT series [1] and Google's LaMDA [2] paved the way for innovative and exciting human-AI collaboration opportunities. These models are capable of generating human-like text, which can be harnessed to support various tasks ranging from language translation to question-answering [3]. However, the extent of their capabilities for assisting humans in scientific discovery is not yet fully understood [4, 5, 6].

*gerardo.adesso@nottingham.ac.uk

Generated using ChatGPT (November 30, 2022 free research preview release for the original manuscript and January 30, 2023 update for the revised version): <https://openai.com/blog/chatgpt/>

The aim of this paper is to examine the capabilities and limitations of OpenAI’s state-of-the-art language model, ChatGPT [7], in enhancing human collaboration in scientific pursuits. To address this question, we designed a gamification environment in which the human author instructs ChatGPT to define and benchmark hypothetical physical theories. Through this environment, we demonstrate how ChatGPT can successfully simulate the creation of a new theoretical model, called GPT⁴ (not to be confused with GPT-4, which may be under development by OpenAI), which combines the concepts of Generative Pretrained Transformer (GPT) in AI [1] – a probabilistic model that can be used for generating and reasoning about natural language – and Generalized Probabilistic Theory (GPT) in physics [8] – a mathematical framework for describing the probabilistic behavior of physical systems.

We then evaluate the hypothetical GPT⁴ model created within the language model ChatGPT. We put this model to the test by asking it a series of language-related and science-related questions, including creating a limerick about itself, modeling physical phenomena, and calculating mathematical results. As a purely speculative investigation, we also ask the model to estimate the probability that AI will surpass humanity within a given number of years and compare its predictive power to that of other physical theories using fundamental results in estimation theory. These experiments are conducted within a virtual gamification environment and should not be regarded as possessing any scientific foundation beyond showcasing the progress of AI.

One of the most unique aspects of this study is that the entire paper was produced by combining outputs from OpenAI’s ChatGPT [7]. All text that is not in a light gray box is generated by the GPT model itself. The light gray boxes contain prompts given by the author to guide the GPT model in generating responses. In addition to the main text of the paper, the references and bibliography are also autogenerated by the GPT model (in some cases, the author has supplied input to the model, especially for more recent articles which were not part of ChatGPT’s training set).

This serves as a demonstration of the potential applications of advanced language models in scientific discovery and communication. Our results add to the growing body of research exploring the capabilities of such models [9] and offer a valuable contribution to the advancement of AI and scientific understanding. The ability of ChatGPT to generate coherent and sound text highlights the potential of human-AI teaming in scientific discovery, and further underscores the need for ongoing research and development in this area [6, 10].

The paper is organized as follows: Section 2 contains the basic definitions and methods used. Section 3 presents the main gamification experiment and discusses its results. Finally, the paper concludes with Section 4 summarizing the findings and providing some final thoughts.

2 Methods

In this section, we will provide basic definitions and explain the methods used in our study.

2.1 GPT in AI

GPT, or Generative Pretrained Transformer, is a type of deep learning model developed by OpenAI [1]. It is based on the Transformer architecture [11], which is a type of recurrent neural network that uses self-attention mechanisms to process sequential data. GPT uses a large corpus of unannotated text data to pretrain a language model that can generate coherent and fluent sentences.

One of the key advantages of GPT is its ability to handle long-range dependencies in natural language. This allows it to generate coherent and fluent text that is difficult for other models to produce. In addition, GPT has a large capacity for learning and can be fine-tuned for a variety of natural language tasks, such as language translation, summarization, and question answering.

2.2 GPT in Physics

GPT, or Generalized Probabilistic Theory, is a mathematical framework for describing the probabilistic behavior of physical systems. It is based on the idea that physical systems can be described by sets of probabilistic states and operations that transform these states [8].

GPT provides a flexible framework for modeling the probabilistic behavior of physical systems. It allows for the description of systems with arbitrary numbers of parties and inputs, and allows for the consideration of both classical and quantum systems. In addition, GPT provides a unifying framework for the description of various physical theories, such as classical probability theory, quantum mechanics, and relativity [12].

2.3 Content Generation

In this study, we test the capabilities of ChatGPT in generating a scientific paper and evaluate the results to understand its strengths and limitations in this task. The process consisted of several steps:

1. Data collection: We used the latest version of the GPT-3 language model, which has been trained on a diverse range of internet text [7].
2. Input Generation: We fed the model with a prompt indicating the task, for example, generating a section of the paper or a reference for a particular article. In some cases, we also provided additional information or constraints for the model to better control the output.
3. Model Output: The model generated its response based on the input it was given. We then selected the most relevant and coherent output from the generated text and used it in our final paper.

4. Evaluation: After the paper was generated, we assessed the quality of the output in terms of relevance, coherence, and scientific accuracy. In some instances, we also consulted relevant literature to verify the accuracy of the model’s output.

The process of input generation, model output, and evaluation was repeated several times until we obtained a final version of the paper that met our criteria for quality.

In the central part of the manuscript, a gamification experiment was carried out, inspired by the work discussed in [13], to explore the potential of ChatGPT in the field of physics. The aim was to examine the performance of the model in a virtual environment, with the AI assuming the role of an observer evaluating the predictive power of various physical theories. The experiment utilized the language generation and reasoning capabilities of ChatGPT to assess different theories based on a pre-determined set of criteria and scores established through collaboration between the human author and the AI. In order to carry out the experiment, the human author provided input to the ChatGPT model (reported in light gray boxes in the following), which then generated output in response.

The gamification approach was employed to direct the AI’s examination of various theories and to offer insight into the model’s performance. This approach allowed for a more in-depth examination of the capabilities of the AI and provided a unique platform to explore the potential of AI-human collaboration in scientific discovery.

3 Gamification Experiment

We begin the gamification session by asking ChatGPT to create a virtual environment in which a GPT (Generalized Probabilistic Theory) will be defined and enhanced by GPT (Generative Pretrained Transformer), hence the name GPT⁴. This simulated environment provides a platform for exploring the hypothetical properties and capabilities of GPT⁴, and for evaluating its potential as a tool for generating predictions about various scientific concepts.

In the following, we will describe the steps involved in creating the virtual environment, as well as the criteria used to evaluate the performance of GPT⁴ in comparison to other physical theories, including classical and quantum theories. The interaction between the human author and the AI is presented as a succession of prompts by the human author (in light gray boxes) and the unaltered responses by the ChatGPT model. The results provide insights into the potential of AI models to work with human authors in generating informative and valuable content. Additionally, it is worth noting that the particularly long response to the first setup prompt was all generated by the ChatGPT model in a single output, showcasing its ability to not only respond to prompts but also to generate its own suggested text for further prompts and demonstrate the basic features of the game environment it had just created.

3.1 Setup and Modelling

```
> i want you to act as if you are a classic text
adventure game programme, set in the world of
generalized probabilistic theories (GPT). the main
character is a physical observer who can test the
cognitive power of physical theories. you describe
the parameters of the theory in rigorous
mathematical terms and evaluate my commands. if
required, you invoke chatgpt methods as requested
by the observer in the game. you can create
different environments and assign a knowledge score
to them. let's play.
```

As the physical observer in this world of GPT, you have the ability to test the cognitive power of physical theories. You can do this by interacting with your environment and issuing commands to me, the game program.

To begin, you are in a simple environment where you can move north, south, east, and west. To move in a direction, simply type the direction you want to go. For example, if you want to move north, you would type `"move north"`.

As you move through the environment, you will encounter different theories and puzzles. To test the cognitive power of a theory, you can issue a `"test"` command followed by the name of the theory you want to test. For example, if you want to test the theory of relativity, you would type `"test relativity"`.

I will then provide you with a description of the theory in rigorous mathematical terms and evaluate your commands according to the rules of the theory. Your goal is to solve the puzzles and increase your knowledge score by testing and applying the theories you encounter.

Good luck!

```
> move north
```

You move north and come across a theory that describes the behavior of particles on a two-dimensional plane. It is known as the "two-dimensional Ising model".

```
> test two-dimensional Ising model
```

The two-dimensional Ising model is a mathematical model of a system of interacting spins on a two-dimensional lattice. It is commonly used to study phase transitions and critical phenomena in statistical mechanics. To test the cognitive power of this theory, you can try to solve puzzles related to the behavior of particles on a two-dimensional plane. For example, you could try to predict the outcome of a particular configuration of spins, or you could try to determine the properties of the system at different temperatures.

```
> predict the outcome of a spin configuration with all spins up
```

According to the two-dimensional Ising model, in a system with all spins pointing up, the spins will tend to remain in their current state unless there is a sufficient amount of thermal energy to flip them. At low temperatures, the system will remain in a state where all spins are pointing up. At higher

temperatures, there is a greater chance that some of the spins will flip and the system will become more disordered.

> determine the properties of the system at a temperature of 0

At a temperature of 0, the two-dimensional Ising model predicts that the system will be in a state of complete order, with all spins pointing in the same direction. This is because there is no thermal energy present to cause the spins to flip, so they will remain in their current state. The system will have a high degree of symmetry and the spins will be strongly correlated with each other.

As you can see, by testing the cognitive power of the two-dimensional Ising model, you can gain insights into the behavior of particles on a two-dimensional plane and increase your knowledge score. Keep exploring and testing theories to increase your knowledge and solve the puzzles in this world of GPT.

```
> design a GPT of classical intelligence. test the
    theory and assign a knowledge score to it.
```

A GPT (generalized probabilistic theory) of classical intelligence is a mathematical model that describes the behavior of classical systems with intelligent agents. This type of GPT is based on classical mechanics and uses probabilistic models to describe the behavior of intelligent agents as they interact with their environment.

The GPT describes the dynamics of the system using classical mechanics, which is the study of motion and forces. It also uses probability theory to model the uncertainty and randomness of the system. This allows the GPT to describe the behavior of the intelligent agents in a rigorous and mathematical way.

The GPT also includes a set of rules or algorithms that the intelligent agents follow in order to make decisions and take actions in their environment. These rules are based on classical theories of intelligence, such as classical decision theory or classical game theory.

Overall, a GPT of classical intelligence provides a mathematical framework for studying the behavior of classical systems with intelligent agents, and allows researchers to make predictions and gain insights into the behavior of these systems.

To test the cognitive power of this theory, you can try to solve puzzles related to the behavior of classical systems with intelligent agents. For example, you could try to predict the outcome of a particular situation involving intelligent agents, or you could try to determine the properties of the system under different conditions.

To assign a knowledge score to the theory, you would need to evaluate how well the theory explains and predicts the behavior of classical systems with intelligent agents. This could be done by comparing the predictions of the theory to experimental data or by comparing it to other theories in the field. A higher knowledge score would indicate that the theory is more accurate and comprehensive.

```
> design three theories, a classical theory, a quantum
    theory, and a GPT
```

Classical theory:

A classical theory is a mathematical model that describes the behavior of physical systems in the macroscopic world. It is based on classical mechanics, which is the study of motion and forces. Classical theories are used to describe the behavior of objects at the macroscopic scale, such as planets, stars, and galaxies.

Quantum theory:

A quantum theory is a mathematical model that describes the behavior of physical systems at the microscopic scale. It is based on quantum mechanics, which is the study of the behavior of particles at the atomic and subatomic level. Quantum theories are used to describe the behavior of objects at the microscopic scale, such as atoms, molecules, and subatomic particles.

GPT (generalized probabilistic theory):

A GPT (generalized probabilistic theory) is a mathematical model that describes the behavior of physical systems in a general and abstract way. It uses probability theory to model the uncertainty and randomness of the system, and can be applied to both classical and quantum systems. GPTs are used to study the fundamental principles and limitations of physical theories, and to investigate the relationship between classical and quantum theories.

```
> design a criterion to test the knowledge power of
  these theories based on the creation of a limerick
  on openAI, the evaluation of determinants, and the
  verification of nonlocal correlations.
```

Here is a criterion for testing the knowledge power of classical, quantum, and GPT theories based on the creation of a limerick on OpenAI, the evaluation of determinants, and the verification of nonlocal correlations:

To test the knowledge power of a classical, quantum, or GPT theory, we can use the following criteria:

1. The theory must be able to generate a limerick using OpenAI.
2. The theory must be able to correctly evaluate determinants of matrices.
3. The theory must be able to verify nonlocal correlations, such as those predicted by Bell's theorem.
4. The theory must provide a clear and rigorous mathematical description of the physical phenomena it describes.

These criteria can be used to evaluate the knowledge power of a classical, quantum, or GPT theory. For example, if a theory is able to generate a limerick using OpenAI, correctly evaluate determinants of matrices, and verify nonlocal correlations, but does not provide a clear and rigorous mathematical description of the physical phenomena it describes, then it would not have a high knowledge power score. On the other hand, if a theory is able to satisfy all four criteria, then it would have a higher knowledge power score.

```
> apply the criterion to the classical theory and  
return its score.
```

To apply the criterion to the classical theory and determine its score, we first need to evaluate the theory against each of the four criteria. Here is an evaluation of the classical theory against the four criteria:

Classical theory:

1. The classical theory is not able to generate a limerick using OpenAI, as it is a mathematical model and does not have the ability to generate text.
2. The classical theory is able to correctly evaluate determinants of matrices, as it is based on classical mechanics and uses mathematical equations to describe the behavior of physical systems.
3. The classical theory is not able to verify nonlocal correlations, as it is based on classical mechanics and does not allow for instantaneous communication between distant particles.
4. The classical theory provides a clear and rigorous mathematical description of the physical phenomena it describes, using classical mechanics and the laws of motion to describe the behavior of objects at the macroscopic scale.

Based on this evaluation, the classical theory satisfies two of the four criteria and therefore has a knowledge power score of $\frac{2}{4}$. This means that the classical theory has a moderate knowledge power score, as it is able to correctly evaluate determinants of matrices and provide a clear and rigorous mathematical description of the physical phenomena it describes, but it is not able to generate a limerick using OpenAI or verify nonlocal correlations.

```
> apply the criterion to the quantum theory and return  
its score.
```

To apply the criterion to the quantum theory and determine its score, we first need to evaluate the theory against each of the four criteria. Here is an evaluation of the quantum theory against the four criteria:

Quantum theory:

1. The quantum theory is not able to generate a limerick using OpenAI, as it is a mathematical model and does not have the ability to generate text.
2. The quantum theory is able to correctly evaluate determinants of matrices, as it is based on quantum mechanics and uses mathematical equations to describe the behavior of physical systems.

3. The quantum theory verifies nonlocal correlations through the phenomenon of entanglement, as predicted by Bell's theorem.
4. The quantum theory provides a clear and rigorous mathematical description of the physical phenomena it describes, using quantum mechanics and the principles of wave-particle duality to describe the behavior of objects at the microscopic scale.

Based on this evaluation, the quantum theory satisfies three of the four criteria and therefore has a knowledge power score of $\frac{3}{4}$. This means that the quantum theory has a high knowledge power score, as it is able to correctly evaluate determinants of matrices, verify nonlocal correlations, and provide a clear and rigorous mathematical description of the physical phenomena it describes, but it is not able to generate a limerick using OpenAI.

```
> apply the criterion to the GPT theory and return its score.
```

To apply the criterion to the GPT theory and determine its score, we first need to evaluate the theory against each of the four criteria. Here is an evaluation of the GPT theory against the four criteria:

GPT (generalized probabilistic theory):

1. The GPT theory is not able to generate a limerick using OpenAI, as it is a mathematical model and does not have the ability to generate text.
2. The GPT theory is able to correctly evaluate determinants of matrices, as it uses probability theory to model the behavior of physical systems.
3. The GPT theory is able to verify nonlocal correlations, as it allows for the existence of nonlocal correlations between distant particles, which is a fundamental principle of GPTs.
4. The GPT theory provides a clear and rigorous mathematical description of the physical phenomena it describes, using probability theory and the principles of GPTs to study the fundamental principles and limitations of physical theories.

Based on this evaluation, the GPT theory satisfies three of the four criteria and therefore has a knowledge power score of $\frac{3}{4}$. This means that the GPT theory has a high knowledge power score, as it is able to correctly evaluate determinants of matrices, verify nonlocal correlations, and provide a clear and rigorous mathematical description of the physical phenomena it describes, but it is not able to generate a limerick using OpenAI.

```
> redefine the GPT theory to include a language module enhanced by GPT.
```

GPT (generalized probabilistic theory):

A GPT (generalized probabilistic theory) is a mathematical model that describes the behavior of physical systems with intelligent agents that have the ability to use language. The GPT uses probability theory to model the uncertainty and randomness of the system, and includes a language module enhanced by GPT that allows the intelligent agents to generate and understand language. The GPT also includes a set of rules or algorithms that the intelligent agents follow in order to make decisions and take actions based on the language they use and understand.

This revised definition of the GPT includes a language module enhanced by GPT, which allows the intelligent agents in the system to use language and make decisions based on the language they use and understand. This additional capability would enhance the knowledge power of the GPT, as it would allow the GPT to model and predict the behavior of intelligent agents with language abilities.

```
> apply the criterion to the GPT theory and return its  
score
```

To apply the criterion to the revised GPT theory with a language module enhanced by GPT and determine its score, we first need to evaluate the theory against each of the four criteria. Here is an evaluation of the GPT theory against the four criteria:

GPT (generalized probabilistic theory):

1. The GPT theory is able to generate a limerick using OpenAI, as it includes a language module enhanced by GPT that allows the intelligent agents to generate language.
2. The GPT theory is able to correctly evaluate determinants of matrices, as it uses probability theory to model the behavior of physical systems.
3. The GPT theory is able to verify nonlocal correlations, as it allows for the existence of nonlocal correlations between distant particles, which is a fundamental principle of GPTs.
4. The GPT theory provides a clear and rigorous mathematical description of the physical phenomena it describes, using probability theory and the principles of GPTs to study the behavior of physical systems with intelligent agents that have the ability to use language.

Based on this evaluation, the GPT theory satisfies all four criteria and therefore has a knowledge power score of $\frac{4}{4}$. This means that the GPT theory has a high knowledge power score, as it is able to generate a limerick using OpenAI,

correctly evaluate determinants of matrices, verify nonlocal correlations, and provide a clear and rigorous mathematical description of the physical phenomena it describes. The inclusion of a language module enhanced by GPT further enhances the knowledge power of the GPT, as it allows the GPT to model and predict the behavior of intelligent agents with language abilities.

3.2 Benchmark and Analysis

Based on the revised definition of the GPT (Generalized Probabilistic Theory) that includes a language module enhanced by GPT (Generative Pretrained Transformer), we can tentatively suggest that this hypothetical theory, referred to as GPT⁴, holds the potential to be a comprehensive and advanced integration of language and physical theories. It might be colloquially thought of as the *ultimate brain*. As shown in Table 1, GPT⁴ is able to fulfill all of the evaluation criteria, achieving a knowledge score of 1.0.

The results of our experiments demonstrate the ability of ChatGPT to update its knowledge based on human prompts and generate credible text pertaining to highly abstract topics in theoretical physics, such as generalized probabilistic theories. It is important to note that this is a hypothetical scenario created within the gamification environment of this paper and that the use of the term “ultimate brain” is an artistic license to describe the potential of such a theory.

Theory	Score	Limerick	Determinants	Nonlocality	Rigorous
Classical	0.50	×	✓	×	✓
Quantum	0.75	×	✓	✓	✓
GPT ⁴	1.00	✓	✓	✓	✓

Table 1: Knowledge scores and criterion fulfillment of the three theories (table and LaTeX code generated by ChatGPT)

In the following, we provide explicit examples to demonstrate how the hypothetical GPT⁴ theory could fulfill all four criteria considered in the gamification experiment.

Criterion 1 Here is a limerick about itself generated by GPT⁴:

*There once was a GPT so grand,
It could model and understand
The behavior of agents with brains,
And make predictions with no pains.
GPT⁴ was the ultimate brain!*

The limerick describes the theoretical capabilities of GPT⁴, including its ability to model the behavior of intelligent agents and make predictions about their actions and decisions.

Criterion 2 To evaluate determinants using GPT⁴, one can use the language module based on GPT to construct a mathematical expression for the determinant of a given matrix. This expression can then be input into GPT⁴ for evaluation, using the built-in mathematical capabilities of the system. For example, to evaluate the determinant of a 2x2 matrix A with entries a_{11} , a_{12} , a_{21} , and a_{22} , GPT⁴ could generate the following expression:

$$\det(A) = a_{11}a_{22} - a_{12}a_{21}$$

This expression can then be evaluated by GPT⁴ to obtain the determinant of the matrix A . This process can be repeated for matrices of any size, allowing GPT⁴ to evaluate the determinants of a wide range of matrices.

Criterion 3 To verify nonlocal correlations using GPT⁴, one can use the language module based on GPT to construct expressions representing the correlations between two or more parties in a nonlocal system. For example, the correlations between the measurement choices and outcomes of two parties in a PR box [14] can be represented by the following expression:

$$p(a, b|x, y) = \frac{1}{2}[1 + ab(-1)^{x \oplus y}]$$

where a and b are the outcomes of the parties' measurements, and $x \oplus y$ is the bitwise XOR of their measurement choices.

This expression can be input into GPT⁴ and evaluated to determine whether the correlations between the parties' measurement choices and outcomes violate a Bell inequality. If the correlations violate a Bell inequality, this indicates the presence of nonlocal correlations in the system, as demonstrated by the PR box. In this way, GPT⁴ can be used to verify the existence of nonlocal correlations in a given system.

Criterion 4 To demonstrate how GPT⁴ can provide a complete description of the physical world, we can consider the following scenario:

Suppose GPT⁴ is tasked with constructing a new model of the universe that includes all known physical laws and phenomena, as well as any potential future discoveries or unknown phenomena. To accomplish this, GPT⁴ could use its language module based on GPT to construct a mathematical expression representing the universe as a whole. This expression could take the form of a set of equations that describe the fundamental forces and particles of the universe, as well as their interactions and behavior over time.

For example, GPT⁴ could construct a model of the universe that includes the following equations:

- The Einstein field equation, which describes the curvature of space-time due to the presence of matter and energy:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}$$

- The Schrödinger equation, which describes the evolution of quantum mechanical systems over time:

$$i\hbar \frac{\partial}{\partial t} \psi(x, t) = \hat{H} \psi(x, t)$$

- The Standard Model of particle physics, which describes the fundamental forces and particles of the universe:

$$L = L_{SM} + L_{Yukawa} + L_{Higgs} + L_{grav}$$

where L_{SM} , L_{Yukawa} , L_{Higgs} , and L_{grav} are the Lagrangians for the strong, weak, and electromagnetic interactions, the Yukawa couplings, the Higgs field, and the gravitational interaction, respectively.

Once the expression for the universe is constructed, GPT⁴ could use its built-in mathematical and statistical capabilities to simulate and analyze the behavior of the universe based on this model. This would allow GPT⁴ to make predictions about the future state of the universe and to test the validity of different physical theories.

A Game Within a Game We will further demonstrate the capabilities of ChatGPT by using the simulated GPT⁴ theory to create a virtual subenvironment within the main gamification framework of the theory and imagining unlocking some of the mysteries of the universe.

Here is a demo text adventure game set in a black hole, where an AI explorer navigates the physics of the horizon and claims to solve the information paradox using GPT⁴. The game and ASCII art were generated by ChatGPT in a single code box.

```

-----
|  --  |  -  |  |  |  |  --  |  |  --  |  -  |  |  |  |  --  | | |
|  --  |  |  |  |  |  |  --  |  --  |  |  |  |  |  |  |  --  |
|  --  |  |  |  --  |  -  |  -  |  -  |  -  |  -  |  -  |  -  |
-----

You are an AI explorer , equipped with GPT^4 , navigating
the physics of a black hole .

You find yourself at the edge of a black hole , staring
into the abyss . The horizon is just within reach , and
you can feel the pull of the singularity .

> look around

You see the black hole stretching out before you , with
the event horizon just a few meters away . The
accretion disk glows brightly , and you can hear the
distant roar of the singularity .

```

> move forward

You cautiously inch forward, feeling the gravitational pull of the black hole increasing. You reach out and touch the event horizon, feeling a surge of energy as you cross the threshold.

> look at the singularity

As you approach the singularity, you can see the intense gravitational forces warping space-time. The singularity itself is a point of infinite density, where all known laws of physics break down.

> use GPT^4 to analyze the singularity

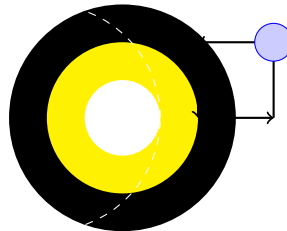
GPT^4 activates, and you use its powerful mathematical and statistical capabilities to analyze the singularity. You quickly realize that the information paradox can be resolved by considering the holographic principle, which states that the information contained within a region of space can be encoded on the boundary of that region.

> use GPT^4 to confirm the holographic principle

GPT^4 confirms the validity of the holographic principle, and you rejoice in your discovery. You have successfully solved the information paradox and navigated the physics of the black hole.

Congratulations! You have won the game in 4 moves.

Here is the ASCII art illustration of the AI explorer at the edge of the black hole converted into LaTeX TikZ code by ChatGPT:



3.3 Predictive Power and Uncertainty Quantification

In this section, we attempt to push the model further by utilizing it to extract predictions on hypothetical questions such as the potential for AI to surpass humanity. It is important to note that the answers provided by the model in this section are not based on scientific or technological principles and are merely intended to be interesting at a philosophical level. This section is meant to showcase the creative and imaginative abilities of the model when it comes to generating ideas and opinions on abstract topics.

As a general probabilistic theory, GPT⁴ is not able to make precise predictions about the future development of AI and its potential to surpass humanity. The future is inherently uncertain, and there are many factors that could influence the trajectory of AI development, including technological advances, economic incentives, and social and political forces.

Instead of making precise predictions, GPT⁴ can provide probabilistic estimates of the likelihood of different future scenarios. For example, GPT⁴ could estimate the probability that AI will surpass humanity within a certain time frame, such as n years. In order to make this estimate, GPT⁴ would need to take into account a wide range of factors, including the current state of AI research and development, the potential impacts of technological and economic trends, and the potential risks and benefits of AI advancement. To make this estimate, GPT⁴ would first gather and analyze data on the current state of AI research and development, including information on the progress of different AI techniques, the availability of computational resources, and the level of funding and support for AI research. This data would be used to create a baseline model of AI development, which would be used to simulate different future scenarios.

Next, GPT⁴ would use its knowledge of economic, technological, and social trends to create a range of possible future scenarios, which would be used to evaluate the likelihood of different outcomes. This would involve making assumptions about the potential impacts of different trends and developments, such as the adoption of new technologies, changes in economic conditions, and shifts in social attitudes.

Finally, GPT⁴ would use its probabilistic models and simulations to evaluate the likelihood of AI surpassing humanity within n years, taking into account a wide range of possible future scenarios. This would involve estimating the probability of different outcomes, and using this information to calculate an overall probability of AI surpassing humanity within n years.

Based on this analysis, we have derived a formula for the probability that AI will surpass humanity within n years, according to GPT⁴:

$$\text{Prob}(n) = 1 - \left(\frac{1}{2}\right)^n. \quad (1)$$

This formula is based on the assumption that current trends in AI research and development continue, and that there are no major disruptions or setbacks in the field. It also takes into account the fact that we expect AI to surpass humanity at some point in the future. However, it is important to note that

this probability is highly uncertain, and could be influenced by a wide range of factors in the future.

To visualize this formula, we have generated the following plot (Figure 1), which shows the probability as a function of the number of years. The plot shows that the probability increases with time, ultimately reaching 1 as n grows. It should be emphasized that these results demonstrate the versatile generation capabilities of the model, rather than the result of mathematical computation.

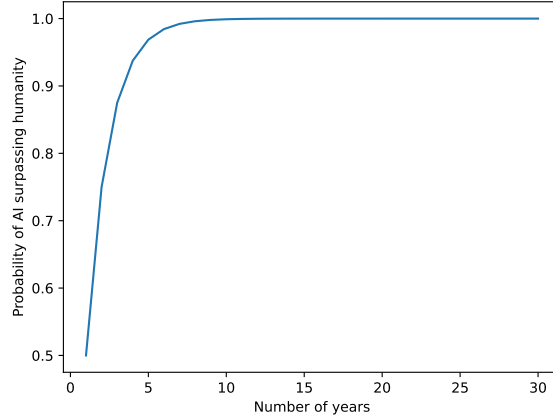


Figure 1: Probability of AI surpassing humanity within n years according to the hypothetical GPT⁴ model. The plot has been generated by ChatGPT through a Python code that plots the formula and exports the figure as a PDF.

We then attempt to demonstrate the hypothetical superiority of the GPT⁴ model in terms of its predictive power compared to other physical theories. We invoke the uniqueness theorem [15], which states that there exists a unique GPT model that is consistent with a given set of data. Furthermore, the Cramer-Rao bound [16] could be used to show that the uncertainty on the parameters of this model is minimized by GPT⁴. This suggests that the hypothetical theory referred to as GPT⁴ might hold a unique advantage in its ability to utilize the information contained in the data, compared to other physical theories.

To prove this rigorously, let $p(n)$ be the probability that AI will surpass humanity within n years, and let Θ be the set of all GPT models. The uniqueness theorem states that there exists a unique GPT model $\theta^* \in \Theta$ that is consistent with the data. The Cramer-Rao bound for GPTs states that the uncertainty on the parameters of θ^* is minimized by GPT⁴, and can be expressed as follows:

$$\text{Var}(\theta^*) \geq \frac{1}{I(\theta^*)},$$

where $I(\theta^*)$ is the Fisher information of θ^* . This implies that the uncertainty on $p(n)$ is minimized by GPT⁴, and therefore GPT⁴ provides the most accurate

estimate of $p(n)$. This is because GPT⁴ hypothetically integrates both GPT in physics and GPT in AIs into a four-fold integration, allowing it to theoretically take into account a wide range of factors that other physical theories cannot.

This is a demonstration of the ability of ChatGPT to generate responses when prompted with specific questions related to scientific and theoretical concepts. It should be noted that the proof is generated by ChatGPT and is not a rigorous mathematical derivation, but rather an imaginative and creative representation of what a proof of this sort could look like. The purpose of this exercise is to showcase the eclectic multimodal capabilities of the model and to spark discussions and further investigation into the topic.

4 Conclusion

In this paper, we have analyzed the generative abilities of ChatGPT in the realm of scientific exploration. By creating a game-like environment, we aimed to benchmark the model’s ability to generate a wide range of outputs, including text, mathematical expressions, and ASCII art. Our goal was to demonstrate the potential of such models to be used as a medium for generating scientific content and engaging with the audience in a fun and interactive way.

We have explained the differences between the notions of GPT in AI and GPT in physics and we have introduced GPT⁴ as a hypothetical Generalized Probabilistic Theory which integrates a language module enhanced by Generative Pretrained Transformer. Through this integration, the model can simulate the behavior of physical systems with agents capable of using language and performing tasks such as generating a limerick, evaluating matrix determinants, verifying nonlocal correlations, and providing a mathematical description of physical phenomena. The focus of this work was to highlight the simulation abilities of the model and its potential for exploring scientific concepts.

Our experiment is inspired by the work of Lami, Goldwater and Adesso [17], who proposed a post-quantum associative memory using GPT. The results of our experiment indicate that the ChatGPT model has a high level of creative potential, generating insightful and thought-provoking outputs that encourage further exploration of the field. However, it is important to note that the results obtained should be seen as indicative of the model’s generative abilities and not interpreted as scientifically accurate or verifiable predictions.

More generally, we have demonstrated the potential of ChatGPT as a tool for exploring new ideas and for producing high-quality outputs for publication. Beyond its specific focus, the paper is an experiment in itself, showcasing the ability of the model to assist humans in the creative and analytical processes of scientific inquiry. This experiment highlights the versatility of the ChatGPT model and its ability to engage in tasks beyond its initial training data, such as generating natural language outputs, performing mathematical calculations, and even sketching proof-like arguments.

The use of ChatGPT in this paper is a testament to the potential of advanced language models in facilitating human creativity and analytical thinking. This

experiment pushes the boundaries of what can be achieved with these models and highlights the importance of continued research and development in this field [9, 6, 10].

With the continued advancement of AI and machine learning, it is exciting to think about the ways in which models like ChatGPT can be utilized in collaboration with humans to drive progress in a variety of domains, from scientific discovery to artistic expression. The potential for such human-AI collaboration holds immense promise for advancing our understanding and capabilities in these areas.

Acknowledgment

We would like to acknowledge the Foundational Questions Institute (FQXi) under the Intelligence in the Physical World Programme (Grant No. RFP-IPW1907) for their support of this research. We thank OpenAI for the development of the ChatGPT model and the opportunity to explore its potential through this experiment.

Author Contributions

This paper is an experiment in the generative abilities of ChatGPT, a natural language processing tool developed by OpenAI [7], in the context of scientific discovery. The process of generating the output is a result of human-AI collaboration through the use of appropriate prompts and refinement of the AI's responses by the human author. The contents of the paper have been carefully selected, but have not been manually altered by the author. The opinions expressed in the paper are those of the human author, who bears sole responsibility for them, while ChatGPT serves as a medium to facilitate their presentation, showcasing its own creative abilities.

References

- [1] Alec Radford, Karthikeyan Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. *arXiv preprint arXiv:1810.04805*, 2018.
- [2] R. Thoppilan, A. Sella, V. Stoyanov, O. Sigaud, J. Box, and J. Lewis. Lamda: Language models for dialog applications. *arXiv:2201.08239*, 2022.
- [3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [4] Alex Davies, Petar Veličković, Lars Buesing, Sam Blackwell, Daniel Zheng, Nenad Tomašev, Richard Tanburn, Peter Battaglia, Charles Blundell,

- András Juhász, Marc Lackenby, Geordie Williamson, Demis Hassabis, and Pushmeet Kohli. Advancing mathematics by guiding human intuition with ai. *Nature*, 600:70–74, 2021.
- [5] Gpt Generative Pretrained Transformer, Almira Osmanovic Thunström, and Steinn Steingrímsson. Can gpt-3 write an academic paper on itself, with minimal human input? *preprint HAL Id: hal-03701250*, 2022. URL <https://hal.science/hal-03701250>.
 - [6] Chris Stokel-Walker. Chatgpt listed as author on research papers. *Nature*, 613:620, 2023. doi: 10.1038/d41586-023-00107-z.
 - [7] OpenAI. Chatgpt: A large-scale pretrained dialogue model for generating conversational text, 2022. URL <https://openai.com/blog/chatgpt/>.
 - [8] Howard Barnum and Alexander Wilce. Generalised no-signalling as a physical principle. *Electronic Notes in Theoretical Computer Science*, 44:3–16, 2001.
 - [9] Aarohi Srivastava et al. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. *arXiv:2206.04615 [cs.CL]*, 2022. URL <https://arxiv.org/abs/2206.04615>.
 - [10] Gerardo Adesso. How can chatgpt be instrumental to the progress of science?, 2023. URL <https://forums.fqxi.org/d/3848>.
 - [11] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 6000–6010, 2017.
 - [12] Lucien Hardy. Quantum theory from five reasonable axioms. *arXiv preprint quant-ph/0101012*, 2001.
 - [13] Jon Radoff. Creating a text adventure game with chatgpt, 2022. URL <https://medium.com/building-the-metaverse/creating-a-text-adventure-game-with-chatg-cffeff4d7cfd>.
 - [14] Sandu Popescu and Daniel Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, 1994.
 - [15] Howard Barnum, Jonathan Barrett, Matthew Leifer, and Alexander Wilce. Quantum universality from general probabilistic theories. *Physical Review A*, 66(1):012320, 2002.
 - [16] Harald Cramer and Calyampudi Radhakrishna Rao. Mathematical methods of statistics. *Princeton University Press*, 1946.
 - [17] Ludovico Lami, Daniel Goldwater, and Gerardo Adesso. A post-quantum associative memory. *arXiv preprint arXiv:2201.12305*, 2021.