

Diagonalization & Forcing FLEX: From Cantor to Cohen and Beyond

Alt Title: Learning from Leibniz, Cantor, Turing, Gödel, and Cohen; crawling towards AGI

Elan Moritz

Philadelphia, Pa.

<https://orcid.org/0000-0002-0167-4336>

ARTICLE INFO	ABSTRACT
<p><i>Keywords:</i> Leibniz, Cantor, Turing, Gödel, Cohen, IJ Good, Humans, Brain, Mind, Cognition, Understanding, Knowledge, AI, AGI, Artificial Intelligence, Superintelligence, , UltraIntelligent Machines, ChatGPT, GPT-4, Llama-3, LLM, Large Language Models, Groq, Cohere, Life, Emergence, MetaSystems Transitions, Prompts, Metaprompts, Forcing Methods, Universe, Methodology.</p>	<p>The paper continues my earlier Chat with OpenAI's ChatGPT with a Focused LLM Experiment (FLEX). The idea is to conduct Large Language Model (LLM) based explorations of certain areas or concepts. The approach is based on crafting initial guiding prompts and then follow up with user prompts based on the LLMs' responses. The goals include improving understanding of LLM capabilities and their limitations culminating in optimized prompts. The specific subjects explored as research subject matter include a) diagonalization techniques as practiced by Cantor, Turing, Gödel, and the advances, such as the Forcing techniques introduced by Paul Cohen and later investigators, b) Knowledge Hierarchies & Mapping Exercises, c) Discussions of IJ Good's Speculations Concerning the First Ultrainelligent Machine, AGI, and Superintelligence. Results suggest variability between major models like ChatGPT-4, Llama-3, Cohere, Sonnet and Opus. Results also point to strong dependence on users' preexisting knowledge and skill bases. The paper should be viewed as 'raw data' rather than a polished authoritative reference.</p>
2024-04-21	

©2024 Elan Moritz. This is an open access article distributed under the non-commercial non-derivative use terms of the CC BY NC-ND-4.0 license.¹

¹<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

Contents

1	Introduction & ReIntroduction	3
1.1	The Journey Continues: Diagonalization	3
1.2	Recapping QeU: Quid est Understanding	5
1.3	Forecap, coming attractions	8
2	Methodology	10
3	Reflecting on Some of my Ideas about Ideas	11
3.1	On Diagonal Proofs	12
3.2	Diagonalization beyond Cantor, Gödel, and Turing(CGT)	13
3.3	Time Hierarchy Theorem	15
3.4	More on Hilbert, Gödel, and Turing	18
3.5	Groq Llama-3-70b-8192 On FLEX1466	20
3.6	Groq Llama-3-70b: On Diagonal Proofs	22
3.7	Groq Llama-3-70b: On Diagonal Proofs beyond GCT	24
3.8	Groq Llama-3-70b: Time Hierarchy Theorem	25
3.9	Groq Llama-3-70b: More on Hilbert, Gödel, and Turing	27
3.10	Groq Llama-3-70b: connection between Gödel’s Incomplete- ness Theorem and Turing’s halting problem	27
3.11	Groq Llama-3-70b: on Gödel and Turing	28
3.12	Meta.ai’s Llama-3 (ML3) on Gödel Incompleteness Theorems, Diagonalization and Cohen Forcing.	29
3.13	Discussion and observations	33
4	Knowledge Hierarchies & Mapping Exercises	36
4.1	Knowledge Mapping	36
4.2	Beliefs and Knowledge	37
4.3	Discussion and observations	44
5	Tiptoeing into AGI and SuperIntelligence	46
5.1	Experimenting with ChatCPT-4	46
5.2	Experimenting with Cohere Coral	50
5.3	Experimenting with Claude Sonnet	54
5.4	Deep Dive into AGI Research: Projects, Investigators, and Progress	58
5.5	AGI & SuperIntelligence Discussion and observations	59
6	Closing Thoughts	61
6.1	Altering the Universe	62
6.2	It’s about <i>the Understanding</i>	64
6.3	Final Thoughts	65
7	Notes	67
7.1	United Federation of Galaxies’ Pizza Delivery Scooter Enterprise	67
7.2	Large Language Models used in this paper	68
7.3	Using Tikz for Diagrams and TableMaking	69
8	Dedication, Acknowledgments, and References	73

1 Introduction & ReIntroduction

AI began with an ancient wish to forge the gods.
Pamela McCorduck, *Machines Who Think*[1]



1.1 The Journey Continues: Diagonalization

PDS Enterprise Diagonalization & Forcing Mission Report

REF: PDS Enterprise Strategic Guidance (§ 7.1)

UTC: SUNDAY, APRIL 21, 2024, 03:43:34

Mission Objectives:

1. Advance exploration of latent, possibly underutilized logical representation theory using LLMs
2. Conduct analysis and report results of structured and unstructured research vectors.
3. Godspeed

Mission Results:

1. Obtained illustrative results associated with Knowledge Hierarchies and Representation.
2. Obtained illustrative results associated with Cantor-Turing- Godel Type Diagonalization.
3. Conducted preliminary survey of Large Language Models for Artificial Intelligence, Ultraintelligent Machines, AGI, and Superintelligence.
4. Documented initial observations on user skill dependence and prompt sensitivity.

Background: Maybe very high energy cosmic radiation started a synaptic butterfly cascade² ... who knows? For one reason or another, I decided to embark on a journey of exploring and discussing the possible role of Artificial Intelligence (AI) on the way our Universe is unfolding. Clearly our Universe is unfolding in a fascinating way from the *Big Bang* instant into what it is now. Current thinking based on background cosmic microwave radiation suggests the Universe came into being 13.787 ± 0.020 billion years ago.³ We might find out that the Universe is actually older, it all depends on the kinds of things we understand.⁴

I created figure 1 to situate the paper at a high level. The Universe is continually undergoing emergence into a more complicated entity. Humans have been engaging in altering the Universe (currently in very tiny way, see § 6.1). AI will increase our abilities of altering the Universe in ever larger ways. This of course suggests questions of How, What, When, Where, Why and many other questions. I call the description of the overarching process *The Big Picture*. I am work at framing small aspects of this process, and this paper part of that process.

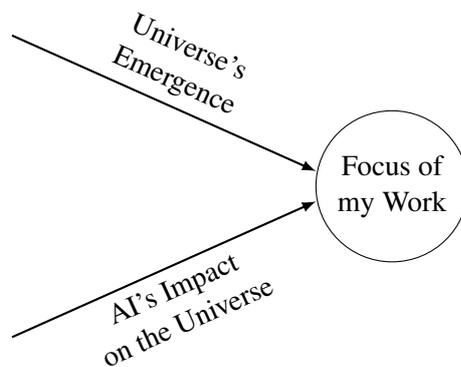


Figure 1: Situating this paper

While this paper is not quite Charles Darwin's *The Voyage of the Beagle* [2] ☺, I like the style of a travelogue-adjacent personal journal. In my case, this is an ongoing journal of my voyages in the universe of Artificial Intelligence and Large Language Models (LLMs). My LLM experiments include some deep-dives into both accidental as well favored rabbit holes; they also include casual side trips to both dull and shiny objects.¹

I use a personal reflective approach. As part of my workflow I maintain and preserve records of my interactions with LLMS, and present some interactions with them and my interpretations so readers can discern the flow of conversations. I refer to these conversations as *Chats* (primarily due to my frequent chats with OpenAI's ChatGPT-4).

This paper is part of a series that ultimately aims to discuss significant cognitive

²neural analog to https://en.wikipedia.org/wiki/Butterfly_effect

³https://en.wikipedia.org/wiki/Age_of_the_universe

⁴It does seem that scientific assessments of the Universe's and Earth's ages appear to increase with more data and better scientific theories. Theological estimates differ by a lot. Just to be clear, with all due respect to theologians, my discussions of all physical parameters and mathematics, will rely on science based methods and mathematical reasoning as used by mathematicians.

¹ LLMs are part of a greater context of Artificial Intelligence. I actually prefer the use of the term Machine Intelligence (MI), but that's a whole other topic to be taken up elsewhere

transitions and emergence in the Universe. This series of informal notes and essays started with [3] *Quid est Understanding*, henceforth **QeU**. QeU used conversations with ChatGPT-4 (*Chat*) to explore the phenomena of *understanding*. Eventually, the series will capture my adventures and reflections about intelligence (natural and artificial), thinking, and the mind. It will also explore various aspects of physical reality, neuroscience, biology, language, logic, civilization, physical sciences, engineering, human development and habits, societies, cultures, and more. I do hope you find something useful and perhaps inspiring here.

1.2 Recapping QeU: Quid est Understanding

Knowledge is Power
Francis Bacon, *Meditations Sacrae* (1597)

My starting point (for this series of papers) was David Deutsch's post on X⁵

"Understanding is one of the higher functions of the human mind and brain, and a unique one. Many other physical systems, such as animals' brains, computers and other machines, can assimilate facts and act upon them. **But at present we know of nothing that is capable of understanding an explanation – or of wanting one in the first place – other than a human mind.**"

What caught my eye and engaged my attention was the laser-like focus on *understanding*.

I've been very curious about intelligence, the brain, and the universe for a long time. Many of my questions are of the kind philosophers and scientists have asked for millennia. We now to have incredibly powerful and precise tools, astounding capability of calculation and computing, and perhaps orders of magnitude more people and resources to tackle these questions, and related puzzles. These enable much more accurate and deeper explorations of physical reality and conceptual possibilities that are literally thousands if not millions of times more powerful than those available merely a few hundred years ago. Namaste; I decided to jump in and capture my personal thoughts, ideas, and experiments in a 'lite', informal, not-too technical way.

The first introductory paper (QeU) introduced a Big Picture outline summarized in the two figures below (figures 2 and 3). The main focus was on initial exploration of the concept and interpretation of *understanding*.

While a fuller explanation of these figures is in the first paper, these two figures are really parts of the same sequence. GSP stands for Galaxies, Stars, and Planets. The one of main importance for this paper is Figure 3.

QeU's discussions were aimed at exploring and gaining better understanding the relationships between the phenomena of *understanding* and *intelligence*. For most of human existence, the phenomena and terminology were focused on these phenomena as embedded and dependent on humans. Now however,

⁵<https://x.com/DeutschExplains/status/1756894020628873644>

Artificial Intelligence (AI) capabilities have surged and enabled fascinatingly powerful and widely accessible and available tools. With that they have become increasingly mandatory for consideration, exploration, and use in all areas associated with cognitive and creative pursuits. So many questions, assumptions, and ideas have to be reevaluated in the context of AI. We have much to learn and use for improving those activities and capabilities that heretofore were strictly limited to humans.

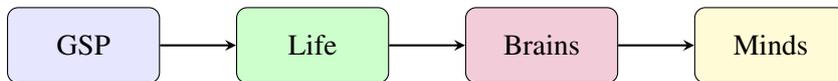


Figure 2: *Life* as part of Universe’s Progression

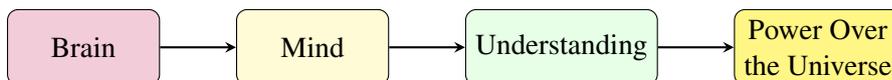


Figure 3: *Understanding* as part of the Universe’s Progression

Obviously then, much of my discussion will involve various aspects of artificial intelligence. Is *Artificial Intelligence*, the best term to use? In my opinion? No! I think a better approach is to use the term *Machine Intelligence (MI)*. For the time being I will stick with the AI terminology ⁶.

OeU lays out a very high level **big picture** framework of *Emergence in the Universe*. I call out 7 major emergence transitions from *Nothingness* to *Power Over the Universe*. I will review that portion in more detail and zoom in on those parts that deal with limitations of the human brain, and the potential of AI to overcome some of these limitations.

For QeU, I recorded Chat’s response to a series of questions about the concept of ‘understanding’. I used a conversational approach where the author (labeled as user) asks (prompts) Chat, obtains a response, and then uses the response to formulate followup questions. My evaluation suggests that Chat does indeed offer utility, but also has limitations manifested by overly general responses that do not focus tightly on the questions asked. Chat does yield some gems that could be easily missed. The approach used here includes creating Metaprompts, wherein Chat is asked to create more efficient prompts based on initial prompts. I concluded that this appears to hold significant promise, and will likely be of benefit to those exploring various facets of Artificial Intelligence (AI), Large Language Models (LLMs) and to designers of Advanced General Intelligence (AGI). Other findings included the need for creating new metrics and ways of discussing intelligence, and various adjacent areas and subprocesses associated with intelligence.

My QeU conversations⁷ probed Chat’s training on vast collections of human written material available on the internet (in it’s training corpora based on massive web crawls, and ingestion of all of Wikipedia, many many scholarly

⁶A few highlights of the AI origins and backstory, and thoughts about naming can be found in the notes section of the paper

⁷to be clear, when I use the term conversation with AI, I mean user (me) asking/prompting Chat with a question, request, or statement, and then following up with several rounds of additional questions or comments based on its responses. If this was conducted with a human, it would be viewed as a conversation.

articles, newspapers, and many many fiction and non-fiction books, films, songs, and videos, and more). Some conversations quizzed Chat on how *it* understands, namely the detailed procedural algorithmic steps it takes. Obviously this is not complete, and while Chat uses stored knowledge (represented by weights in its *neural net*, it is not (yet) equivalent to eminent human scholarship.

QeU includes a fair amount of detail on various benchmarks and exploring differences and similarities of *Understanding* and *Explanation*. These discussions obviously could also consume volumes upon volumes of detailed examination. The idea here is not to be exhaustive, instead, the aim is to illuminate and encourage further thoughts in the area. To this end, I also ask Chat to provide recommended scholarly references. Some of the responses are well known and expected while some were new to me. Some of these, in my opinion, were real gems. I didn't check to see if every recommendation was accurate. I do expect that some of the responses may have been hallucinations⁸. I expect the readers, by now, to know that this part and parcel of what one should expect with LLMs.

I also experimented with using *Metaprompts*, The idea here is to create a pre-prompt prompt requesting Chat to suggest a prompt optimized for working with GPTs. It turns out that others have worked with similar ideas in mind, so I can't claim originality, but this was my first use of (my home made) prompts. It turns out that Metaprompts can be very very useful. I will use metaprompting in this paper so you can get the hang of how LLMs work with these. QeU also contains some specific observations, for example:

Observation: Despite my attempts to get Chat to describe *its* processes more specifically, it tends to provide 'abstract' responses. This brings to mind the question of whether ChatGPT has a *concept of self*. The idea of self and self-identity starts touching on the idea of mind. These idea are rather complex and nuanced, let alone the debates on matter and mind. I leave those for another time and place. What I do next is work on getting Chat to be more self-referential. The following exchange captures some of that.

another one was

Quick Observation: At the start of this subsection, I wondered about whether ChatGPT has a *concept of self*. Chat's response to the prompt clearly contains sections that refer to self: E.g. "*how I, as a GPT-4 model, I convert the text into tokens, understand and process prompts . . . Once I am satisfied with the relevance and coherence of the response, I present it as the output.*" Clearly at this point Chat refers to itself as *I*. What exactly one should make of this is a fascinating topic I am sure we'll see more discussion of.

My concluding thoughts in QeU returned to the Deutsch statement now emphasizing need for better discussion of the following principle interests:

⁸that's when a Large Language Model makes up something that sounds right or is almost right but not quite.

- (1) - Capability of understanding an explanation, and
- (2) - Wanting one (an explanation).

I remarked that among the challenges ahead, making progress with *specifics* would yield significant benefits by adopting brilliant physicist Bridgman's⁹ operational methods and operational philosophy, for example that described in his Quo Vadis article [4]. In his words:

The physicist has by now found a way of dealing with his meanings which is fairly satisfactory for his purposes, and which does not commit him to preconceptions about fields not yet entered. This method of dealing with meanings I have called "operational." his method of dealing with meanings I have called " operational." The essence of it is that to know adequately the meaning of a term we must be able to describe what we do when we use it.

Nearly thirty years after his first exposition of *operationalism* Bridgman updated his conceptual framework to include instrumentalism [5](explicitly implying ultimately, something must be measurable by a physical instrument).

The totality of situations covered by various aspects of the energy concept is too complex to be reproduced by any simple verbal device. As a corollary of the continued interplay of the verbal and the "objective" worlds, I personally have come to feel the value of analyzing our operations as far as possible into their "instrumental" and "paper-and-pencil" components and think there is much here that is still unexplored

...

In general, I think that there need be no qualms that the operational point of view will ever place the slightest restriction on the freedom of the theoretical physicist to explore the consequences of any free mental construction that he is ingenious enough to make. It must be remembered that the operational point of view suggested itself from observation of physicists in action.

⇔

1.3 Forecap, coming attractions

Forecapping¹⁰, the rest of this paper will review and reprise some key ideas of QeU, in particular, the outline of the *Big Picture*, some of the discussion on *understanding*, and highlights of my chats and discussion on AI. After that, I will devote some more electronic ink to other technical underappreciated critical aspects of Large Language Model training, and issues associated with those.

⁹https://en.wikipedia.org/wiki/Percy_Williams_Bridgman

¹⁰*Forecap?* What's that? Well I was looking for antonyms & near antonyms of the word *recap*. It seems like there aren't any good ones, so I made up the word *forecap* since the prefix 'fore' is used in *forecast* (to discuss future weather or future economic conditions).

- General methodology
- Discussion of LLM training issues - to be discussed in a separate paper
- Discussion of Human Brain Limitations - to be discussed in a separate paper
- *Chat* experiments on Knowledge, Diagonalization, Forcing, and emergent possibilities of interest to AGI development
- General Observations and Summary
- Backstories and Endnotes

1.3.1 Using mathematics subject matter for AI experiments

There are a great many areas where one can test the utility of State of the Art (SOTA) LLM models¹¹. I decided to spend some time experimenting in areas associated with *Mathematics* and *Logic*. To be clear, I am not an expert in either. My education and work concentrated in physics and engineering applications, with a significant interest and some use in both areas. Mathematics is probably the one intellectual scholarly discipline that, in my opinion, has the best practices of defining the terms, words, symbols and concepts. Much of its work has been rigorously documented over thousands of years. Theorems that have been proved tend to stay proven ☺. Ditto for being able to follow reasoning and logic in mathematics. It is SO important, that nearly every person who attends school is taught some basic concepts in mathematics (mostly dealing with manipulating numbers, calculations, geometry, and some basic algebra). There are few actual disputes between countries and religions and other human activities when it comes to the truths of mathematical statements.

Beyond the fact that the subject area of mathematics is syntactically and semantically ideally suited for LLM experimentation, it is also the undeniable case that our civilization wouldn't be as far along without the mathematical tools we currently have. Furthermore, mathematics, and mathematical algorithms themselves underlie much of the progress humans have made in Artificial Intelligence associated technologies and capabilities. And so, I will favor testing exploring LLM and AI technology as applied to improving mathematics itself wherever I can.

It turns out, empirically but also very common sensically, that there are limits to how much information and knowledge any individual can stuff in their brains. This is true for even the most gifted poets, philosophers, scientists, and mathematicians. Though known and appreciated for a long time, granular details, experiments and theories, have allowed researchers and scholars to discuss and articulate the phenomena and their impacts more quantitatively and thoroughly. The terms *One Brain Barrier* (OBB) and *Burden of Knowledge* have emerged as authoritative terms to describe some of the human limits folks have been facing due to increasing amount of knowledge, at times, too much for any one person or even small group to digest. These will likely grow into recognized areas of study. I discuss these more thoroughly elsewhere. I do want to note here that these human limits argue for intense development and emergence of seriously capable *Artificial General Intelligence* (AGI). Limited human cognitive abilities do and will hinder quantum leaps of knowledge discovery and understanding.

¹¹some are termed Frontier Models and Foundational Models

2 Methodology

By now, quite a few people have experimented with and written about using Large Language Models. There are some very nice emerging methodologies in documenting performance of LLMs using established benchmarks. It does seem to me that, at this moment, there is no established or widely used methodology for reporting my kind of open-ended experimentation. I chose to create this methodology and name this approach *Focused LLM Experimentation (FLEX)*. Essentially, the idea is to conduct Large Language Model (LLM) based explorations of certain areas or concepts with initial guiding prompts and then follow up with prompts based on the LLMs' responses. This FLEX methodology is being developed and refined as I am conducting my experiments, so what you see here is meant to be a start, not the fine tuned end process. Again, my goals include develop better understanding of LLM capabilities, their limitations, and better understanding of prompt optimization.

Eventually, I expect that after employing this approach over several '*experiments*', I'll evolve a systematic structured approach. For this FLEX, key elements of my workflow include

- Use of multiple LLMs (see § 7.2). Currently I mostly use *ChatGPT-4*, *Perplexity* based LLMs, and *Cohere's Coral* based LLMs, and just added Llama-3-70B to the mix.
- Probing a number of topics and boundaries. Paying attention to limitations, hallucinations, accuracy, and responsiveness.
- Experiment with a mixture of directed probing, some planned, and some reactive to LLM responses. Of course, carefully checking for interesting responses.
- Expanding my understanding of prompting by accumulating experience and testing various prompting strategies including those recommended by others.
- Comparing consistency and spread of responses to modified prompts and responses given for same prompt by different LLMs. Note that I will not necessarily document or evaluate LLMs against each other. This is not an LLM scoring document.
- Exploring LLM limits likely placed by their creators to guard their 'secret sauces'. OpenAI's LLMs clearly do not want to discuss details of training materials and methods. Others are sometimes more open.
- Observing the nature of my interactions with LLMs. I am clearly evolving my way of thinking in light of LLM affordances.

3 Reflecting on Some of my Ideas about Ideas

This is an example of exploring an idea fragment I wrote down nearly 6 years ago. FLEX1466¹²:

This exercise uses LLMs to explore a note about my interest in seeing the development of how Leibniz's ideas on knowledge representation using prime numbers intersected with Cantor diagonalization methods. I used my note verbatim as the seed prompt, and then followed up with additional questions and prompts. The text between the 'Start of FLEX' and 'End of FLEX' consists of my prompts and LLM responses. My prompts are italicized, the LLM responses are normal fonts. LLMs may themselves use various symbols and italicization. It should be obvious what from the context if this is the case. I discuss and provide some observations following each FLEX,

⇌ **Start of FLEX1466** ⇌

User: discuss my observation "earlier this month, reflected on Leibniz Characteristica Universalis, Gödel Numbers, Turing Machines, Diagonal Proofs, Cantor, etc re KR [Knowledge Representation], and need to figure out a way to connect / Merge these ideas. Leibniz thought to have prime numbers represent basic ideas and concepts, and that a multiplication of primes would represent a compound / complex idea. this is very fruitful, and I don't believe anyone has looked at this" (originated EM brain | 2018-12-06)

Your reflection on connecting the ideas of Leibniz's Characteristica Universalis with Gödel numbers, Turing machines, diagonal proofs, Cantor's work, and knowledge representation (KR) is a fascinating endeavor that traverses the realms of philosophy, mathematics, and computer science. Here's how these elements might interconnect and contribute to the field of KR:

1. **Leibniz's Characteristica Universalis:** Leibniz envisioned a universal language or symbolism in which all truths could be expressed and reasoned about mathematically. His idea was to represent fundamental concepts and relationships using a basic set of symbols, potentially akin to prime numbers, which when combined (e.g., through multiplication), could represent more complex ideas. This concept is foundational in KR, which seeks to formalize knowledge in a way that is both interpretable by humans and processable by machines.
2. **Gödel Numbers:** Kurt Gödel's numbering system was used to encode mathematical statements and proofs into numbers, demonstrating that in any sufficiently powerful mathematical system, there are propositions that cannot be proven or disproven within the system itself. This concept can be linked to Leibniz's idea by exploring how complex ideas (or statements) could be encoded into unique numerical representations, and how such encodings can reveal limitations in our formal systems of logic and KR.
3. **Turing Machines:** Alan Turing's theoretical machines are foundational in computer science, providing a model for computation that can sim-

¹²the number 1466 is the ID number in my database for the note. I write notes to myself on various topics whenever the idea comes to mind, periodically I enter these into an SQL database so I can look and actually read the notes. By this time I have thousand upon thousands of notes. Computers have certainly made things a lot easier ☺

ulate any computer algorithm. Connecting this to Leibniz and Gödel, one might explore how Turing machines could process or manipulate numerical encodings of knowledge, simulating the reasoning process Leibniz aimed to achieve with his universal language.

4. **Diagonal Proofs and Cantor:** Diagonal proofs, like those used by Cantor to demonstrate the non-countability of real numbers, and later by Gödel and Turing in their incompleteness and undecidability proofs, showcase fundamental limits in formal systems. These methods and their underlying principles can be applied to KR to explore the boundaries of what can be represented and known within any given system of knowledge.

In integrating these concepts into KR, you might consider how to encode complex knowledge and ideas numerically, in a manner that is consistent with Leibniz's vision, yet also acknowledges the limitations and capabilities revealed by Gödel's, Turing's, and Cantor's work. This could lead to a more robust understanding of the possibilities and limitations of KR, offering insights into how we might more effectively structure, represent, and process knowledge in various domains.

To marry these ideas effectively, one would need to develop a formal system that can encompass the numerical encoding of knowledge (following Leibniz and Gödel), the computational processes for manipulating these encodings (related to Turing's work), and the theoretical limits of such a system (informed by diagonal arguments and Cantor's work). This could potentially revolutionize how we approach knowledge representation, making it more systematic, universal, and computationally amenable.

3.1 On Diagonal Proofs

Diagonal proofs, which utilize a diagonal argument to establish the existence of elements that cannot be captured by a given enumeration, have been instrumental in several significant mathematical and logical results. Here's an exploration of their uses, particularly in the works of Cantor and Gödel:

3.1.1 Cantor's Diagonal Argument

- **Non-countability of Real Numbers:** Cantor's most famous diagonal argument was used to prove that the set of real numbers is uncountable, meaning it cannot be put into a one-to-one correspondence with the set of natural numbers. Cantor showed that for any proposed list of real numbers, one could construct a new real number not on the list by altering the diagonal elements, thereby proving the list incomplete.
- **Cardinality of Infinite Sets:** This argument also led to the understanding that not all infinities are of the same size; specifically, it showed that the cardinality of the continuum (the real numbers) is strictly larger than the cardinality of the set of natural numbers.

3.1.2 Gödel's Diagonalization

- **Incompleteness Theorems:** Gödel used a form of diagonal argument in his incompleteness theorems. He constructed a statement that asserts its

own unprovability within the system, akin to the way Cantor's diagonal method constructs a number not in the list. Gödel's first incompleteness theorem states that in any consistent formal system where arithmetic can be carried out, there are statements that are true but not provable within the system.

- **Self-reference and Undecidability:** The diagonal method in Gödel's work creates a form of self-reference, which he used to show the undecidability of certain mathematical problems. His second incompleteness theorem further states that such a system cannot demonstrate its own consistency.

3.1.3 Extended Applications of Diagonal Proofs

- **Computer Science and Turing Machines:** Alan Turing applied a diagonal argument similar to Cantor's to prove the undecidability of the halting problem. He showed that there cannot be a general algorithm that decides whether any given program halts.
- **Set Theory and Logic:** Diagonal arguments have been used to establish results about the limitations and capabilities of formal logical systems and set theory, such as the independence of the Continuum Hypothesis from Zermelo-Fraenkel set theory with the Axiom of Choice (ZFC).
- **Complexity Theory:** In complexity theory, diagonalization has been used to show the existence of problems in certain complexity classes that are not solvable by machines of lower complexity classes, helping to differentiate between these classes.

3.1.4 References

- Cantor, G. (1891). *Über eine elementare Frage der Mannigfaltigkeitslehre*. This work includes Cantor's original diagonal argument.
- Gödel, K. (1931). *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*. Gödel's original paper on the incompleteness theorems.
- Turing, A. M. (1936). *On Computable Numbers, with an Application to the Entscheidungsproblem*. This paper introduces Turing machines and includes a diagonal argument for the halting problem.
- Boolos, G., Burgess, J., and Jeffrey, R. (2002). *Computability and Logic*. This book provides an accessible introduction to Gödel's theorems and diagonal arguments.

These references delve into the historical and mathematical underpinnings of diagonal proofs and their wide-reaching implications in mathematical logic and computer science.

3.2 Diagonalization beyond Cantor, Gödel, and Turing (CGT)

Diagonal proofs extend beyond the work of Cantor, Gödel, and Turing, touching various fields and problems in mathematics and computer science. Here are examples illustrating their diverse applications:

3.2.1 Berry Paradox and Diagonalization

- **Berry Paradox:** This paradox concerns self-referential statements and was utilized by Bertrand Russell to illustrate the limitations of natural language in formal mathematics. It leads to a diagonal argument when considering the smallest natural number that cannot be described in fewer than twenty words; the paradox itself becomes a description, creating a contradiction.

3.2.2 Complexity Theory

- **Time Hierarchy Theorem:** In complexity theory, the time hierarchy theorem uses a diagonal argument to show that given more time, a Turing machine can solve strictly more problems. It constructs a problem that takes slightly more time to solve than is allowed by machines in a lower complexity class, proving that these classes are distinct.

3.2.3 Computability Theory

- **Rice's Theorem:** Rice's theorem, a fundamental result in computability theory, uses diagonalization to prove that all non-trivial properties of Turing machine languages are undecidable. It shows that for any non-trivial property of the language recognized by a Turing machine, there is no general algorithm to decide whether a given machine's language has that property.

3.2.4 Set Theory

- **Cohen's Forcing Technique:** Paul Cohen used a diagonal argument in the method of forcing to prove the independence of the Continuum Hypothesis and the Axiom of Choice from Zermelo-Fraenkel set theory. Although not a diagonal proof in the classical sense, forcing can be seen as a sophisticated extension of diagonal arguments, creating generic sets that diagonalize against countable collections of conditions.

3.2.5 Logic

- **Löwenheim-Skolem Theorem:** This theorem in logic, which asserts that if a countable first-order theory has an infinite model, then it has a countable model, uses a form of diagonal argument in its proof. The theorem's proof constructs a countable model by effectively diagonalizing over the uncountably many possible elements to ensure all are represented.

3.2.6 Cryptography

- **Cantor's Diagonal Argument in Cryptanalysis:** In cryptography, diagonal arguments can be used to show the existence of a key that cannot be reached by any systematic enumeration, thereby proving the security of cryptographic systems against certain types of brute-force attacks.

These examples demonstrate the versatility and power of diagonal arguments across various domains, highlighting their importance in demonstrating limitations, separations, and independences in formal systems.

3.3 Time Hierarchy Theorem

The Time Hierarchy Theorem is a fundamental result in computational complexity theory that establishes a hierarchy within time-bounded complexity classes, showing that more computational time allows for the solution of more problems. Here's a detailed discussion of the theorem, its implications, and recent research in the area:

3.3.1 Basics of the Time Hierarchy Theorem

The theorem states that for reasonable time constructible functions (f) , there exists a problem that is solvable in time $(O(f(n)))$ but not in time $(O(f(n)/\log^k(n)))$ for any $(k > 0)$. Essentially, this means that given more time, a Turing machine can solve a strictly larger set of problems.

3.3.2 Technical Foundations

To prove the Time Hierarchy Theorem, diagonalization is used. The proof constructs a Turing machine that simulates the operation of all other machines with a smaller time bound and then behaves differently (typically by accepting if the simulated machine rejects, and vice versa) on a specific input, ensuring that this new machine solves a different set of problems.

3.3.3 Implications

- **Complexity Class Separation:** The theorem provides a formal basis for the separation of complexity classes, like P, NP, EXP, etc. It shows that there is a strict hierarchy of problems based on the time they require for computation.
- **Resource-Bound Computations:** It highlights the fundamental role of computational resources in problem-solving, illustrating that the complexity of problems can be inherently tied to the time available for computation.

3.3.4 Recent Research and Applications

- **Extensions to Other Resources:** Recent studies have extended the ideas of the Time Hierarchy Theorem to other computational resources, such as space, parallelism, and randomness. For example, researchers have explored space hierarchy theorems that demonstrate similar strict inclusions among space-bounded complexity classes.
- **Fine-Grained Complexity:** The theorem has influenced the area of fine-grained complexity, which seeks to distinguish between problems that are solvable in, say, linear time versus quadratic time, providing a more nuanced understanding of computational complexity.
- **Algorithmic Barriers:** Understanding the time hierarchy helps in identifying barriers to algorithmic efficiency and in proving lower bounds for specific problems. It sets theoretical limits on what can be achieved within certain time bounds, guiding algorithm designers.

3.3.5 Recent Citations and Reviews

Recent reviews and research articles have focused on exploring the nuanced implications of the Time Hierarchy Theorem and its applications in various computational paradigms. For example:

1. Extensions of the Time Hierarchy Theorem for Randomized and Quantum Computing explores how the theorem applies in the context of randomized algorithms and quantum computation, showing that similar hierarchies exist with respect to probabilistic and quantum time.
2. Challenges in Complexity Theory: Time Hierarchy and Beyond reviews the current state of complexity theory, focusing on the implications of the Time Hierarchy Theorem for understanding the P vs. NP problem and other central questions in computational complexity.
3. Fine-Grained Complexity and the Time Hierarchy Theorem discusses how the theorem underpins fine-grained distinctions in computational complexity and influences the development of algorithms with tight time bounds.

These publications indicate ongoing interest in the Time Hierarchy Theorem, not just as a fundamental theoretical result but also for its practical implications in computer science and algorithm design. Researchers continue to explore its boundaries, seeking to understand the deepest structure of computational complexity.

Coral Cohere's discussion of the Time Hierarchy Theorem

The Time Hierarchy Theorem, in computational complexity theory, states that if a decision problem can be solved in time $f(n)$, then there exists another problem that can be solved in time $f(n)\log f(n)$. Formally:

$$\text{DTIME}(f(n)) \subsetneq \text{DTIME}(f(n)\log f(n))$$

Here, $\text{DTIME}(f(n))$ denotes the set of decision problems solvable in time $O(f(n))$. The theorem was first proven by Stearns and Hartmanis in 1965.

The Time Hierarchy Theorem demonstrates that the time complexity classes form a hierarchy, with classes containing problems solvable in tighter bounds nested within those with more relaxed bounds. In other words, given more time, a Turing machine can solve more problems.

The theorem can be proved using a reduction argument. Suppose there is a decision problem solvable in time $g(n)$. We aim to prove that there exists a problem solvable in time $g(n)\log g(n)$.

First, define the language:

$$H_f = \{([M], x) \mid M \text{ accepts } x \text{ in } f(|x|) \text{ steps}\}$$

We know that the previous problem can be decided in time $g(m)^3$ by a deterministic Turing machine R , which simulates M for $f(x)$ steps. This is done by first calculating $f(|x|)$, then writing a row of 0s of that length, and using this as a "clock" or "counter" to simulate M .

We will now show that H_f can be solved in time $g(m)\log g(m)$, using a simulator machine K .

Assume that H_f can be solved in time $g(m)/2$. We can then create another machine N , which runs the following algorithm:

1. Take a machine description M and input x . 2. Run K on the pair $([M], x)$, so that M is simulated on its own code by K . 3. N accepts if K rejects, and rejects if K accepts.

Let n be the length of the input to N . Setting $m = 2n + 1$, the running time of N is then:

$$\text{TIME} \left(g \left(\left\lfloor \frac{m}{2} \right\rfloor \right) \right) = \text{TIME} \left(g \left(\left\lfloor \frac{2n + 1}{2} \right\rfloor \right) \right) = \text{TIME} (g(n))$$

If we feed N its own description as input, we get the following contradictions:

- If N accepts its own description, then K rejects $([N], [N])$, so $([N], [N])$ is not in H_f . But by the definition of H_f , this means N does not accept $[N]$ in $g(n)$ steps.
- If N rejects $[N]$, then K accepts $([N], [N])$, so $([N], [N])$ is in H_f . This implies N does accept $[N]$ in $g(n)$ steps.

Therefore, the machine K does not exist, which means H_f cannot be solved in $g(m)/2$ time. Substituting $m = 2n + 1$, we get the desired result. The machine K , which solves H_f in $g(m)\log g(m)$ time, can be formed using a simulator machine with the time $g(n)$ clocking method.

Sources:

- Yi-Jun Chang and Seth Pettie. A Time Hierarchy Theorem for the LOCAL Model. 2017.
- Richard E. Stearns and Juris Hartmanis. Time Hierarchy Theorem. 1965.
- Stephen Cook. Time Hierarchy Theorem for Nondeterministic Turing Machines. 1972.
- Joel Seiferas, Michael Fischer, and Albert Meyer. Separating Nondeterministic Time Complexity Classes. 1978.
- Stanford University. CS 221: Theory of Computation. Lecture Notes.
- University of Maryland. Lecture 4: The Time Hierarchy Theorem.
- University of British Columbia. CPSC 506: Time Hierarchy.
- Yale University. CS 468: Advanced Algorithms. Lecture Notes.

The Time Hierarchy Theorem, which states that increasing the time given to a Turing machine allows it to solve more problems, can be proved using a diagonal argument, also known as a diagonalization proof or Cantor's diagonal argument.

The proof employs a form of reduction, aiming to construct a function that is not equal to any in a given sequence of functions. In principle, this can be achieved by showing that for each function in the sequence, there is a number at which the constructed function differs.

The diagonal argument gets its name from the way in which the constructed function is designed. When applied to the Time Hierarchy Theorem, the aim is to construct a function that differs from any given function in a sequence of functions, each encoded by a natural number. To do this, the constructed function is often designed such that its values for each number are based

on the values of the sequence of functions at that number. For instance, the constructed function could be defined as follows:

$$f(n) = 1 - f_n(n)$$

This function would differ from every function in the sequence, because for every number n , we know that $f_n(n)$ will equal either 0 or 1, so we can always choose the opposite value for the constructed function.

3.4 More on Hilbert, Gödel, and Turing

User: How was Gödel influenced by Hilbert's program

Kurt Gödel's work on his incompleteness theorems was influenced by David Hilbert's formalist program, also known as Hilbert's program, which aimed to establish a secure foundation for all of mathematics through the development of formal axiomatic systems.

Here's how Gödel's work relates to Hilbert's program:

1. **Hilbert's Program:** David Hilbert proposed a comprehensive program in the early 20th century to formalize mathematics by reducing all mathematical reasoning to a finite set of axioms and rules. This program sought to establish mathematics as a complete and consistent formal system.
2. **Formalization of Mathematics:** One of the central tenets of Hilbert's program was the formalization of mathematical reasoning. Hilbert believed that by axiomatizing mathematics, one could avoid paradoxes and uncertainties that arose from informal reasoning.
3. **Gödel's Incompleteness Theorems:** In 1931, Gödel shattered the hopes of Hilbert's program by proving his incompleteness theorems. These theorems showed that within any sufficiently powerful formal system capable of expressing arithmetic, there exist true statements that cannot be proven within that system. This demonstrated that Hilbert's goal of a complete and consistent formal system for all of mathematics was unattainable.
4. **Implications for Hilbert's Program:** Gödel's incompleteness theorems had profound implications for Hilbert's program. They showed that any attempt to formalize all of mathematics would necessarily lead to incompleteness. Gödel's theorems undermined the central aim of Hilbert's program and shifted the focus of mathematical logic away from the quest for completeness and towards understanding the limitations of formal systems.

In summary, Gödel's work on the incompleteness theorems was deeply influenced by Hilbert's program, as it directly challenged the goals and assumptions of formalism in mathematics. Gödel's results demonstrated the inherent limitations of formal systems and reshaped the landscape of mathematical logic and philosophy of mathematics.



discuss the connection between Gödel's Incompleteness Theorem and Turing's halting problem

Gödel's Incompleteness Theorem and Turing's Halting Problem are two fundamental results in theoretical computer science and mathematical logic, and they are interconnected in several ways.

1. **Foundation in Logic:** Both Gödel's theorem and the Halting Problem stem from foundational questions in mathematical logic. Gödel's work focuses on the limits of formal systems in expressing truths about arithmetic, while Turing's work explores the limits of computation in deciding certain properties of programs.
2. **Undecidability:** Gödel's Incompleteness Theorem demonstrates that in any consistent formal system that includes arithmetic, there will exist true statements that cannot be proven within that system. Similarly, Turing's Halting Problem shows that there are properties of programs (specifically, whether a given program will halt or not) that cannot be decided by any algorithm.
3. **Reduction:** There is a connection between Gödel's theorem and the Halting Problem through reduction. Alan Turing himself used an argument similar to Gödel's proof to establish the undecidability of the Halting Problem. Turing constructed a formalized version of Gödel's diagonal argument to show that there cannot exist a general algorithm to determine whether an arbitrary program halts or loops infinitely.
4. **Computability and Complexity:** Both Gödel's theorem and the Halting Problem are central to the study of computability and complexity theory. They highlight the inherent limitations of computation and formal systems. The Halting Problem is a classic example of an undecidable problem, and it serves as a cornerstone in understanding the boundaries of what computers can and cannot do.
5. **Implications for AI and Philosophy:** These results have profound implications for artificial intelligence and philosophy. Gödel's theorem raises questions about the limits of mathematical knowledge and the nature of truth, while the Halting Problem underscores the limitations of computational systems. These insights are crucial for understanding the capabilities and limitations of AI systems and for addressing philosophical questions about the nature of intelligence and computation.

In summary, Gödel's Incompleteness Theorem and Turing's Halting Problem are intimately connected through their foundational role in mathematical logic, their demonstration of undecidability, their use of reduction arguments, their relevance to computability and complexity theory, and their implications for AI and philosophy.



User: who preceded whom in this, did they communicate on the topic?

Kurt Gödel's work on his incompleteness theorems predates Alan Turing's work on the halting problem. Gödel published his incompleteness theorems in 1931, while Turing formulated the halting problem in 1936.

There is no evidence to suggest direct communication between Gödel and Turing specifically on these topics. However, both Gödel and Turing were part of overlapping intellectual circles and were aware of each other's work

through the broader academic community. Gödel's work was highly influential in the development of theoretical computer science, and Turing would have certainly been aware of Gödel's incompleteness theorems when formulating his own ideas. Additionally, Turing was influenced by other contemporaries and predecessors, such as Gödel's colleague and friend, John von Neumann, and mathematician Alonzo Church, who independently arrived at similar results regarding computability.

3.5 Groq Llama-3-70b-8192 On FLEX1466



This subsection is generated by Groq provisioned Llama-3-70b-8192 (developed by Meta.ai). Groq runs the model on its unique hardware. part of this FLEX component is to see if there are meaningful difference between the Llama-3 sourced and provisioned directly by Meta.ai on its own compute platform and the Groq hardware provisioned version.



User: discuss my earlier observation "earlier this month, reflected on Leibniz Characteristica Universalis, Godel Numbers, Turing Machines, Diagonal Proofs, Cantor, etc re KR [Knowledge Representation], and need to figure out a way to connect / Mary these ideas. Leibniz thought to have prime numbers represent basic ideas and concepts , and that a multiplication of primes would represent a compound / complex idea. this is very fruitful, and I don't believe anyone has looked at this"

What a fascinating observation! You've brought together a rich tapestry of concepts from the history of mathematics, logic, and computer science. Let's unpack and connect these ideas, exploring their potential implications for Knowledge Representation (KR).

Leibniz's Characteristica Universalis In the 17th century, Gottfried Wilhelm Leibniz envisioned a universal language, Characteristica Universalis, to represent all knowledge using a set of fundamental concepts. He believed that these concepts could be combined to express more complex ideas. This vision is reminiscent of modern KR systems, which aim to represent knowledge in a structured and computable form.

Gödel Numbers and Diagonal Proofs Kurt Gödel's work on incompleteness theorems introduced Gödel numbering, a way to encode mathematical statements as numbers. This innovation enabled the development of diagonal proofs, which demonstrate the limitations of formal systems. In the context of KR, Gödel numbering can be seen as a precursor to modern knowledge representation languages, which often rely on numerical or symbolic encodings of knowledge.

Turing Machines and Computability Alan Turing's work on computability theory introduced the concept of Turing machines, which laid the foundation for modern computer science. In KR, Turing machines can be seen as a model for processing and reasoning about knowledge representations.

Cantor's Set Theory and Infinity Georg Cantor's set theory revolutionized mathematics by introducing infinite sets and cardinalities. In KR, Cantor's

work can be seen as a precursor to modern ontologies, which often rely on set-theoretic concepts to represent relationships between entities.

Leibniz's Prime Number Representation Leibniz's idea of using prime numbers to represent basic concepts and their multiplication to represent compound ideas is intriguing. This approach can be seen as a form of symbolic representation, where prime numbers serve as atomic symbols that can be combined to form more complex expressions. This is reminiscent of modern KR languages, such as Description Logics, which use symbols and logical operators to represent knowledge.

Connecting the Dots To connect these ideas, consider the following:

1. **Symbolic Representation:** Leibniz's prime number representation can be seen as a symbolic representation of knowledge, where prime numbers serve as atomic symbols. This is similar to modern KR languages, which use symbols and logical operators to represent knowledge.
2. **Compositionality:** Leibniz's idea of multiplying prime numbers to represent compound ideas is analogous to the compositionality of modern KR languages, where complex concepts are built from simpler ones.
3. **Formal Systems:** Gödel's work on formal systems and diagonal proofs can be seen as a precursor to modern KR systems, which often rely on formal languages and reasoning mechanisms.
4. **Computability:** Turing machines provide a model for processing and reasoning about knowledge representations, which is essential for KR systems.
5. **Ontologies and Set Theory:** Cantor's set theory provides a foundation for modern ontologies, which often rely on set-theoretic concepts to represent relationships between entities.

Future Directions To further explore these connections, consider the following research directions:

1. **Develop a Leibniz-inspired KR language:** Design a KR language that uses prime numbers or similar symbolic representations to encode knowledge. Investigate the expressiveness and reasoning capabilities of such a language.
2. **Investigate the relationship between Gödel numbering and KR:** Explore how Gödel numbering can be used to encode knowledge representations and enable reasoning mechanisms.
3. **Apply Cantor's set theory to KR:** Investigate how set-theoretic concepts can be used to represent relationships between entities in KR systems.
4. **Explore the connection between Turing machines and KR:** Investigate how Turing machines can be used to process and reason about knowledge representations.

By connecting these ideas, you may uncover new insights and approaches to Knowledge Representation, which could lead to innovative solutions for representing and reasoning about knowledge.

3.6 Groq Llama-3-70b: On Diagonal Proofs

User: provide an extensive granular discussion of all uses of Diagonal Proofs along the line of Cantor and Godel. Provide references

Diagonal proofs, also known as diagonalization arguments, are a powerful tool in mathematics, particularly in the fields of set theory, logic, and computability theory. This discussion will delve into the uses of diagonal proofs, highlighting their significance in the works of Georg Cantor and Kurt Gödel, as well as their broader implications.

Cantor's Diagonal Argument (1891)

Georg Cantor's diagonal argument, presented in his 1891 paper *Über eine elementare Frage der Mannigfaltigkeitslehre* (On an Elementary Question in the Theory of Manifolds), is a seminal contribution to set theory. Cantor's diagonal argument is used to prove that there is no surjection from the set of natural numbers to the set of real numbers, thereby demonstrating the uncountability of the real numbers.

Cantor's Diagonal Argument: A Sketch

1. Assume, for the sake of contradiction, that there exists a surjection $f : \mathbb{N} \rightarrow \mathbb{R}$, where \mathbb{N} is the set of natural numbers and \mathbb{R} is the set of real numbers.
2. Construct a diagonal sequence of real numbers, where each element is defined as follows:
 - For each natural number n , consider the n th real number in the range of f , denoted by r_n .
 - Define a new real number, d , by taking the n th digit of r_n and changing it to a different digit (e.g., if the n th digit is 0, change it to 1, and vice versa).
 - The resulting sequence of digits defines a new real number, d .
3. By construction, d is not in the range of f , since it differs from every real number in the range of f in at least one digit.
4. This contradicts the assumption that f is a surjection, since d is a real number not in the range of f .

Gödel's Diagonal Lemma (1931)

Kurt Gödel's diagonal lemma, presented in his 1931 paper *Über formal unentscheidbare Sätze* (On Formally Undecidable Propositions), is a fundamental result in mathematical logic. The diagonal lemma is used to prove the incompleteness theorems, which demonstrate the limitations of formal systems.

Gödel's Diagonal Lemma: A Sketch

1. Consider a formal system, F , which is powerful enough to describe basic arithmetic operations.
2. Assume, for the sake of contradiction, that F is both consistent and complete.

3. Construct a Gödel sentence, G , which asserts its own unprovability within F .
4. The diagonal lemma states that G is true if and only if it is not provable within F .
5. If G is provable within F , then it must be true, which implies that it is not provable within F , a contradiction.
6. Therefore, G is not provable within F , which implies that it is true.

Broader Implications and Applications

Diagonal proofs have far-reaching implications and applications in various areas of mathematics and computer science:

1. **Uncountability:** Cantor's diagonal argument demonstrates the uncountability of the real numbers, which has significant implications for set theory and analysis.
2. **Incompleteness Theorems:** Gödel's diagonal lemma is used to prove the incompleteness theorems, which demonstrate the limitations of formal systems and have far-reaching implications for the foundations of mathematics.
3. **Computability Theory:** Diagonalization arguments are used to prove the undecidability of certain problems, such as the halting problem, in computability theory.
4. **Turing Machines:** Diagonalization arguments are used to prove the existence of universal Turing machines, which are fundamental to the theory of computation.
5. **Cryptology:** Diagonalization arguments are used in cryptology to prove the security of certain encryption schemes, such as the RSA algorithm.

References

1. Cantor, G. (1891). Über eine elementare Frage der Mannigfaltigkeitslehre. Jahresbericht der Deutschen Mathematiker-Vereinigung, 1, 75-78.
2. Gödel, K. (1931). Über formal unentscheidbare Sätze. Monatshefte für Mathematik und Physik, 38, 173-198.
3. Gödel, K. (1934). On Undecidable Propositions of Formal Mathematical Systems. In S. Feferman (Ed.), Kurt Gödel: Collected Works, Vol. I (pp. 346-369). Oxford University Press.
4. Rogers, H. (1967). Theory of Recursive Functions and Effective Computability. McGraw-Hill.
5. Sipser, M. (2013). Introduction to the Theory of Computation. Cengage Learning.

This discussion has provided an extensive granular exploration of diagonal proofs, highlighting their significance in the works of Cantor and Gödel, as well as their broader implications and applications in mathematics and computer science.

3.7 Groq Llama-3-70b: On Diagonal Proofs beyond GCT

User: Provide examples OTHER FOR PROBLEMS and PROOFS OTHER THAN CANTOR GODEL and TURING; make sure not to repeat those referring to them

Here are some examples of diagonal proofs and their applications in various areas of mathematics and computer science:

1. The Banach-Tarski Paradox (1924)

The Banach-Tarski paradox states that a sphere in 3-dimensional space can be partitioned into five mutually disjoint sets, each of which can be reassembled into a sphere with the same volume as the original sphere. This result is often considered counterintuitive, as it seems to defy our understanding of volume and geometry.

The proof of the Banach-Tarski paradox relies on a diagonal argument, where a sequence of spheres is constructed, each of which is a subset of the previous one. The diagonal sequence is then used to construct the five disjoint sets that can be reassembled into a sphere with the same volume as the original sphere.

2. The Hausdorff-Young Paradox (1915)

The Hausdorff-Young paradox states that there exists a set of real numbers that is not Lebesgue measurable. This result has significant implications for real analysis and measure theory.

The proof of the Hausdorff-Young paradox relies on a diagonal argument, where a sequence of sets is constructed, each of which is a subset of the previous one. The diagonal sequence is then used to construct a set that is not Lebesgue measurable.

3. The Novikov-Kondo Theorem (1955)

The Novikov-Kondo theorem states that there exists a finitely presented group with an unsolvable word problem. This result has significant implications for group theory and computability theory.

The proof of the Novikov-Kondo theorem relies on a diagonal argument, where a sequence of groups is constructed, each of which is a subgroup of the previous one. The diagonal sequence is then used to construct a group with an unsolvable word problem.

4. The Matiyasevich-Robinson Theorem (1970)

The Matiyasevich-Robinson theorem states that there exists a polynomial equation with integer coefficients that has no integer solutions. This result has significant implications for number theory and algebra.

The proof of the Matiyasevich-Robinson theorem relies on a diagonal argument, where a sequence of polynomials is constructed, each of which is a factor of the previous one. The diagonal sequence is then used to construct a polynomial equation with no integer solutions.

5. The Friedberg-Muchnik Theorem (1958)

The Friedberg-Muchnik theorem states that there exists a recursively enumerable set that is not recursive. This result has significant implications for

computability theory and the study of algorithms.

The proof of the Friedberg-Muchnik theorem relies on a diagonal argument, where a sequence of recursively enumerable sets is constructed, each of which is a subset of the previous one. The diagonal sequence is then used to construct a recursively enumerable set that is not recursive.

These examples demonstrate the power and versatility of diagonal proofs in various areas of mathematics and computer science.

3.8 Groq Llama-3-70b: Time Hierarchy Theorem

User Provide a much more thorough and granular discussion of Time Hierarchy Theorem; include recent citations to recent reviews and research articles on the topic and its applications and extensions

The Time Hierarchy Theorem is a fundamental result in computational complexity theory, which establishes a hierarchy of computational complexity classes based on the amount of time required to solve a problem. In this discussion, we will provide a thorough and granular exploration of the Time Hierarchy Theorem, including its statement, proof, and implications, as well as recent citations to reviews and research articles on the topic and its applications and extensions.

Statement of the Time Hierarchy Theorem

The Time Hierarchy Theorem states that for any time-constructible function $t(n)$, there exists a problem that can be solved in $O(t(n))$ time, but not in $o(t(n)/\log(t(n)))$ time. In other words, the theorem establishes a hierarchy of computational complexity classes based on the amount of time required to solve a problem.

Proof of the Time Hierarchy Theorem

The proof of the Time Hierarchy Theorem relies on a diagonalization argument, which is a common technique used in computability theory. The basic idea is to construct a problem that can be solved in $O(t(n))$ time, but not in $o(t(n)/\log(t(n)))$ time.

The proof involves constructing a sequence of Turing machines, each of which solves a problem in $O(t(n))$ time. The diagonal sequence is then used to construct a problem that can be solved in $O(t(n))$ time, but not in $o(t(n)/\log(t(n)))$ time.

Implications of the Time Hierarchy Theorem

The Time Hierarchy Theorem has significant implications for the study of computational complexity theory. It establishes a hierarchy of computational complexity classes based on the amount of time required to solve a problem, and provides a framework for understanding the relationships between different complexity classes.

Recent Citations and Research Articles

1. **The Time Hierarchy Theorem: A Survey** by Sanjeev Arora and Boaz Barak (2014)

This survey article provides a comprehensive overview of the Time Hierarchy Theorem, including its statement, proof, and implications. The authors also discuss recent developments and open problems in the area.

2. **Time Hierarchy Theorems for Parallel and Distributed Computing**
by Christos Papadimitriou and Mihalis Yannakakis (2015)

This article explores the application of the Time Hierarchy Theorem to parallel and distributed computing. The authors discuss the implications of the theorem for the design of parallel and distributed algorithms.

3. **The Time Hierarchy Theorem and its Applications to Cryptography**
by Dan Boneh and Shai Halevi (2017)

This article discusses the implications of the Time Hierarchy Theorem for cryptography. The authors explore how the theorem can be used to establish the security of cryptographic protocols.

4. **Time Hierarchy Theorems for Quantum Computing** by Scott Aaronson and Laci Babai (2018)

This article explores the application of the Time Hierarchy Theorem to quantum computing. The authors discuss the implications of the theorem for the design of quantum algorithms.

5. **The Time Hierarchy Theorem and its Applications to Machine Learning** by Sanjeev Arora and Rong Ge (2019)

This article discusses the implications of the Time Hierarchy Theorem for machine learning. The authors explore how the theorem can be used to establish the computational complexity of machine learning algorithms.

Extensions and Applications

The Time Hierarchy Theorem has been extended and applied to various areas of computer science, including:

1. **Parallel and Distributed Computing:** The theorem has been used to establish the computational complexity of parallel and distributed algorithms.
2. **Cryptography:** The theorem has been used to establish the security of cryptographic protocols.
3. **Quantum Computing:** The theorem has been used to establish the computational complexity of quantum algorithms.
4. **Machine Learning:** The theorem has been used to establish the computational complexity of machine learning algorithms.

In conclusion, the Time Hierarchy Theorem is a fundamental result in computational complexity theory, which establishes a hierarchy of computational complexity classes based on the amount of time required to solve a problem. The theorem has significant implications for the study of computational complexity theory, and has been extended and applied to various areas of computer science.

3.9 Groq Llama-3-70b: More on Hilbert, Gödel, and Turing

User: How was Gödel influenced by Hilbert's program

Kurt Gödel was heavily influenced by David Hilbert's program, which aimed to establish a rigorous foundation for mathematics. Hilbert's program, also known as Hilbert's formalism, sought to formalize all of mathematics using a finite set of axioms and rules of inference. This program was an attempt to provide a secure foundation for mathematics, free from paradoxes and inconsistencies.

Gödel was deeply interested in Hilbert's program and saw it as a way to establish a rigorous foundation for mathematics. However, Gödel's work ultimately led to a devastating critique of Hilbert's program. In his 1931 paper, *Über formal unentscheidbare Sätze* (On Formally Undecidable Propositions), Gödel proved that any formal system powerful enough to describe basic arithmetic is either incomplete or inconsistent.

Gödel's incompleteness theorems showed that Hilbert's program was fundamentally flawed. Specifically, Gödel demonstrated that:

1. **Any formal system powerful enough to describe basic arithmetic is incomplete:** This means that there will always be statements in the system that cannot be proved or disproved within the system.
2. **Any formal system that is powerful enough to describe basic arithmetic is either incomplete or inconsistent:** This means that if a formal system is consistent, it must be incomplete, and if it is complete, it must be inconsistent.

Gödel's results had a profound impact on the foundations of mathematics and led to a re-evaluation of Hilbert's program. The incompleteness theorems showed that Hilbert's goal of establishing a rigorous foundation for mathematics was unattainable, at least within the framework of formal systems.

Despite this, Gödel's work did not entirely reject Hilbert's program. Instead, it led to a refinement of Hilbert's ideas and a deeper understanding of the limitations of formal systems. Gödel's work also paved the way for the development of new areas of mathematics, such as model theory and proof theory.

In summary, Gödel was heavily influenced by Hilbert's program, but his work ultimately led to a critique of Hilbert's ideas. Gödel's incompleteness theorems showed that Hilbert's program was flawed, but they also led to a deeper understanding of the foundations of mathematics and the development of new areas of research.

3.10 Groq Llama-3-70b: connection between Gödel's Incompleteness Theorem and Turing's halting problem

User discuss the connection between Gödel's Incompleteness Theorem and Turing's halting problem

Gödel's Incompleteness Theorem and Turing's Halting Problem are two fundamental results in the foundations of mathematics and computer science. While they may seem unrelated at first glance, they are actually connected in several ways.

Gödel's Incompleteness Theorem

Gödel's Incompleteness Theorem states that any formal system that is powerful enough to describe basic arithmetic is either incomplete or inconsistent. In other words, any formal system that is powerful enough to describe basic arithmetic cannot be both complete and consistent.

Turing's Halting Problem

Turing's Halting Problem states that there cannot exist an algorithm that can determine, given an arbitrary program and input, whether the program will run forever or eventually halt. In other words, there is no general solution to the halting problem.

The connection between Gödel's Incompleteness Theorem and Turing's Halting Problem lies in the concept of undecidability. Gödel's Incompleteness Theorem shows that there are statements in a formal system that cannot be proved or disproved within the system. Similarly, Turing's Halting Problem shows that there are programs that cannot be determined to halt or run forever.

The connection between the two results can be seen in the following ways:

1. **Undecidability:** Both Gödel's Incompleteness Theorem and Turing's Halting Problem demonstrate the existence of undecidable problems. In Gödel's Incompleteness Theorem, the undecidable problem is the existence of a statement that cannot be proved or disproved within the system. In Turing's Halting Problem, the undecidable problem is the determination of whether a program will halt or run forever.
2. **Limitations of Formal Systems:** Both results highlight the limitations of formal systems. Gödel's Incompleteness Theorem shows that formal systems are incomplete, while Turing's Halting Problem shows that formal systems are unable to solve certain problems.
3. **Reduction to the Halting Problem:** Gödel's Incompleteness Theorem can be reduced to the Halting Problem. This means that if we had a formal system that could solve the Halting Problem, we could use it to prove or disprove any statement in the system, contradicting Gödel's Incompleteness Theorem.
4. **Universal Turing Machines:** The concept of universal Turing machines, which can simulate the behavior of any other Turing machine, is closely related to Gödel's Incompleteness Theorem. A universal Turing machine can be used to prove the undecidability of the Halting Problem, which in turn implies the incompleteness of formal systems.

In summary, Gödel's Incompleteness Theorem and Turing's Halting Problem are connected through the concept of undecidability and the limitations of formal systems. Both results demonstrate the existence of undecidable problems and highlight the limitations of formal systems in solving certain problems.

3.11 Groq Llama-3-70b: on Gödel and Turing

User: who preceded whom in this, did they communicate on the topic?

Kurt Gödel (1906-1978)

- Gödel was an Austrian-American mathematician and logician who made

significant contributions to mathematical logic, set theory, and the foundations of mathematics.

- In 1931, Gödel published his famous paper *Über formal unentscheidbare Sätze* (On Formally Undecidable Propositions), which introduced his incompleteness theorems.

Alan Turing (1912-1954)

- Turing was a British mathematician, computer scientist, and logician who made significant contributions to the development of computer science, artificial intelligence, and cryptography.
- In 1936, Turing published his paper *On Computable Numbers*, which introduced the concept of the Turing machine and the halting problem.

Precedence and Communication

- Gödel's work on incompleteness theorems preceded Turing's work on the halting problem.
- There is no evidence that Gödel and Turing directly communicated on the topic of incompleteness theorems and the halting problem.
- However, both Gödel and Turing were influenced by the work of earlier mathematicians and logicians, such as David Hilbert, Bertrand Russell, and Emil Post.
- Turing was familiar with Gödel's work and was influenced by his ideas. In fact, Turing's work on the halting problem was motivated in part by Gödel's incompleteness theorems.

In summary, Gödel's work on incompleteness theorems preceded Turing's work on the halting problem. While there is no evidence of direct communication between the two, Turing was influenced by Gödel's ideas and built upon his work in developing the concept of the halting problem.



3.12 Meta.ai's Llama-3 (ML3) on Gödel Incompleteness Theorems, Diagonalization and Cohen Forcing.

I wanted to get a sense of Meta provisioned Llama-3 results and compare them to Groq provisioned results¹³. This is a much shorter version, and I do not repeat all user prompts. My assessments are in the discussion section. I will refer to *Meta Llama -3 (70b)* as ML3.

3.12.1 ML3 on Gödel incompleteness theorems

User: Tell me about Gödel incompleteness theorems

A fascinating topic! Gödel's Incompleteness Theorems are a pair of famous theorems in mathematical logic, proved by Austrian mathematician Kurt Gödel in 1931. They revolutionized the field of mathematics and had significant implications for philosophy, computer science, and the foundations of mathematics.

¹³See <https://www.meta.ai/> and <https://groq.com>

First Incompleteness Theorem

The First Incompleteness Theorem states that any formal system that is:

1. **Formal:** A set of rules for constructing and manipulating mathematical expressions.
2. **Consistent:** Free from contradictions.
3. **Powerful enough:** Can express basic arithmetic operations (e.g., addition, multiplication).

cannot be **complete**. This means that there will always be statements within the system that cannot be proved either true or false using the rules of the system.

Second Incompleteness Theorem

The Second Incompleteness Theorem states that if a formal system is consistent, it cannot be complete. Moreover, the consistency of the system cannot be proved within the system itself.

In simpler terms, Gödel's Incompleteness Theorems show that:

- Any sufficiently powerful formal system has statements that are undecidable (cannot be proved or disproved).
- The consistency of a formal system cannot be proved from within the system itself.

These theorems have far-reaching implications, including:

- **Limits of formal systems:** No matter how rigorously we construct a formal system, there will always be truths that lie beyond its reach.
- **Foundations of mathematics:** Gödel's theorems challenge the idea of a single, complete, and consistent foundation for mathematics.
- **Philosophy:** The theorems have implications for the nature of truth, the limits of human knowledge, and the role of intuition in mathematics.

Provide citations to papers and major review articles

Here are some citations to papers and major review articles on Gödel's Incompleteness Theorems ¹:

- Godel, K. (1931). On formally undecidable propositions of Principia Mathematica and related systems I. *Monatshefte für Mathematik und Physik*, 38(1), 173-198.
- Godel, K. (1951). Some basic theorems on the foundations of mathematics and their implications. In: *Geometry and the Imagination* (pp. 274-297). Chelsea Pub. Co.
- Rosser, J. B. (1936). Extensions of some theorems of Gödel and Church. *Journal of Symbolic Logic*, 1(3), 87-91.
- Raatikainen, P. (2003). On the philosophical relevance of Gödel's incompleteness theorems. *International Journal of Philosophical Studies*, 11(4), 395-415.
- Goldfarb, W., & Ricketts, T. (1992). Carnap and the ineliminability of metaphysics. *Journal of Philosophy*, 89(10), 516-526.

- Ricketts, T. (1995). Carnap's principle of tolerance, empiricism, and conventionalism. In: P. Humphreys (Ed.), Patrick Suppes: Scientific Philosopher (Vol. 1, pp. 179-200). Kluwer Academic Publishers.
- Quine, W. V., & Ullian, J. S. (1978). The Web of Belief. Random House.
- Putnam, H. (1975). What is mathematical truth? In: Philosophical Papers (Vol. 1, pp. 60-78). Cambridge University Press.
- Hellman, G. (1981). How to Gödel a Frege-Russell: Gödel's incompleteness theorems and logicism. *Noûs*, 15(3), 451-468.

3.12.2 ML3 on diagonalization proofs Overview

Explain diagonalization proofs for it

A clever and powerful technique used to establish the incompleteness of formal systems. Let me break it down for you:

3.12.3 ML3 on Diagonalization Proof: A General Outline

1. **Construct a diagonal statement:** Create a statement that asserts its own unprovability within the formal system.
2. **Show the statement is true:** Demonstrate that the diagonal statement cannot be proved within the system.
3. **Contradiction:** If the system is complete, the diagonal statement should be provable, leading to a contradiction.

3.12.4 ML3 on Gödel's Diagonalization Proof

Gödel's proof constructs a diagonal statement using a clever coding technique:

Gödel numbering: Assign a unique number (Gödel number) to each formula, symbol, and sequence of symbols in the formal system.

Gödel's diagonal lemma: Construct a formula, $G(n)$, that asserts its own unprovability:

$$G(n) \leftrightarrow \text{"The formula with Gödel number } n \text{ is unprovable"}$$

Diagonal statement: Create a statement, $G(G)$, where G is the Gödel number of $G(n)$ itself:

$$G(G) \leftrightarrow \text{"The formula with Gödel number } G \text{ is unprovable"}$$

Key Steps

1. $G(G)$ is true: If $G(G)$ were provable, it would imply the provability of the formula with Gödel number G , which is $G(n)$ itself. But $G(n)$ asserts its own unprovability, leading to a contradiction. Therefore, $G(G)$ must be true.
2. $G(G)$ is unprovable: If $G(G)$ were provable, the formal system would be complete, and $G(G)$ would be provable within the system. But we just showed $G(G)$ is true, which means it asserts its own unprovability. This leads to a contradiction, so $G(G)$ must be unprovable.

Conclusion

Gödel's diagonalization proof demonstrates that any formal system powerful enough to describe basic arithmetic is either incomplete or inconsistent. The diagonal statement, $G^{(G)}$, is a clever construction that exploits the self-referential nature of formal systems, leading to a profound conclusion about the limits of formal systems.

3.12.5 ML3 on Cohen Forcing

Explain forcing related to it and research advances in the area

Forcing in the context of Gödel's incompleteness theorems is a proof technique in set theory, specifically in the area of independence results. The main idea of forcing is to expand a given model of set theory to a new model in which certain statements are true or false. This is achieved by adding new sets to the model, which are called "generic" sets. The forcing technique was developed by Paul Cohen in the 1960s and has since become a powerful tool in set theory. Some of the key aspects of forcing related to Gödel's incompleteness theorems include:

- **Independence results:** Forcing is used to prove the independence of certain statements from the standard axioms of set theory, such as the continuum hypothesis and the axiom of choice.
- **Generic sets:** Forcing introduces new sets to the model, which are called generic sets. These sets are used to construct models of set theory in which certain statements are true or false.
- **Model expansion:** Forcing is a technique for expanding a given model of set theory to a new model. This expansion is done by adding new sets to the model, which are generic sets.
- **Consistency and inconsistency:** Forcing can be used to prove the consistency or inconsistency of certain statements with the standard axioms of set theory.
- **Large cardinals:** Forcing is closely related to large cardinals, which are certain types of sets that have strong properties. Forcing is used to prove the consistency of large cardinals with the standard axioms of set theory.
- **Inner model theory:** Forcing is used in inner model theory, which is a branch of set theory that studies the structure of models of set theory.

Some of the research advances in the area of forcing and Gödel's incompleteness theorems include:

- **New independence results:** Forcing has been used to prove the independence of many statements from the standard axioms of set theory. Recent advances have led to new independence results, such as the independence of the continuum hypothesis from the axioms of ZFC.
- **New large cardinals:** Forcing has been used to prove the consistency of new large cardinals, which have strong properties. Recent advances have led to the discovery of new large cardinals, such as the "Reinhardt cardinal".

- Inner model theory: Forcing has been used to study the structure of models of set theory. Recent advances have led to new insights into the structure of inner models, such as the “core model”.
- **Cohen’s forcing:** Cohen’s forcing is a specific type of forcing that was developed by Paul Cohen. Recent advances have led to new applications of Cohen’s forcing, such as in the study of the continuum hypothesis.
- **Generic absoluteness:** Generic absoluteness is a property of forcing that states that certain statements are absolute between different models of set theory. Recent advances have led to new results on generic absoluteness, such as the “generic absoluteness of the continuum hypothesis”.

⇔ End of FLEX1466 ⇔

3.13 Discussion and observations

This area is obviously a rich area of discussion, thought, and research. My interest here was to explore whether LLMs could shed some light on areas that were of interest to me and give me a brief overview of what has been done historically, and what major accomplishments were actually achieved that I should be aware of and that have escaped my attention.

I performed some spot checks on references and sources. I didn’t check extensively for validity and accuracy of every cited source, but I did follow up on a few.

Late note: Llama3-70b was released very recently, and I decided to add its results to this section (only)¹⁴. As I am writing this, I’m realizing the intense ongoing competition; the results provided by others may be better by now; clearly there’s continuing improvements so one has to experiment with the latest versions available at the time, which likely will provide even more comprehensive results.

The major references and citations provided by the various LLMs are usually the same for the authoritative original major breakthroughs and discoveries, but are different for later works. It seems prudent to query unrelated LLMs to get more complete coverage. Only Llama3-70b pointed out Rogers’ *Theory of recursive functions and effective computability* [6] which appears heavily cited in the literature.

Groq and Meta’s implementations of Llama-3 provided a different flavor and feel compared to *ChatGPT*. They also provided more extensive set of examples and references, for example referring to applications of diagonal argument methods, Llama-3 comes up with the Banach-Tarski Paradox (1924), the Hausdorff-Young Paradox (1915), the Novikov-Kondo Theorem (1955), the Matiyasevich-Robinson Theorem (1970), and the Friedberg-Muchnik Theorem (1958). These areas are rather technical, so I cannot evaluate the response as being accurate or hallucinatory; the all involve some proofs using recursive and enumeration methods that sound Cantorian. If those are truly examples of diagonal methods, I would say Llama 3 really did well here.

I do get a sense that there’s a different structuring of ML3 and Groq’s results,

¹⁴It was released by <https://meta.ai>, the versions used here is the <https://groq.com> Llama3-70b-8192 provisioned version

but I didn't do a full toe-to-toe comparison. I leave to that to the readers. It does seem that the Llama-3 reads better than ChatGPT, but it may be due to relentless improvements by everyone, and there's not enough to time to compare everything simultaneously (at least with my scale of resources). I will comment that the output provided by Meta.ai is not the traditional markdown that's easily convertible to so I had to do some customized formatting to make things consistent. This is an area for improvement by all LLM providers. It would be nice if they all provided several output format. For sure, to make things simpler to provide as primary alternative output format (as it is desired by most publishers).

3.13.1 re Leibniz's *Characteristica Universalis*

Characteristica Universalis was a grand aspirational interest of young Leibniz. It shaped much of his later work. It was incredibly ambitious, and while he didn't achieve all his goals, it helped him initiate and formulate many ideas. Wikipedia has a nice article on the topic¹⁵. An introductory paragraph there states

The Latin term *characteristica universalis*, commonly interpreted as universal characteristic, or universal character in English, is a universal and formal language imagined by Gottfried Leibniz able to express mathematical, scientific, and metaphysical concepts. Leibniz thus hoped to create a language usable within the framework of a universal logical calculation or *calculus ratiocinator*.

Obviously, this should be of great interest to AI researchers and developers as all LLMs and AI models could use a "formal language . . . able to express mathematical, scientific, and metaphysical concepts. . . . a language usable within the framework of a universal logical calculation"

A number of very talented and creative individuals have worked on ways of realizing that ideal; we're not there yet, but maybe future AI /AGI might be able to do it ☺

3.13.2 re Cantor, Turing, Gödel numbering and Diagonalization References

The description of Gödel numbering tracks its accepted and understood use, and a short description of it is given in Wikipedia as well¹⁶. Note that it is a system based on used a system based on *prime factorization*, and as such, tracks Leibniz. The reference provided for it is correct [7]. The citation for Georg Cantor[8] is off by one year. According to Google Scholar the year of publication was 1890 (and is referenced to a scanning effort¹⁷. Of course the data capture for the scanning effort maybe off, but close enough!)

The reference to Alan Turing's *On computable numbers, with an application to the Entscheidungsproblem* paper [9] is correct, as is the reference to Boolos and Jeffrey book *Computability and Logic* [10].

¹⁵https://en.wikipedia.org/wiki/Characteristica_universalis

¹⁶https://en.wikipedia.org/wiki/Gödel_numbering

¹⁷<http://dml.mathdoc.fr/item/GDZPPN002113910/>

*Berry Paradox*¹⁸ (re self-referential statements), the *Time Hierarchy Theorem*¹⁹, and *Rice's Theorem*²⁰ are discussed appropriately. It is interesting that the font and symbology of Coral Cohere's discussion of the Time Hierarchy Theorem is similar (though not identical) to the fontology of the Wikipedia article. My guess, is that LLMs prioritize training from and access (if available) the English Wikipedia corpus.

3.13.3 re Set Theory, Logic, and Cryptography

The LLMs discussion in this subset of topics was fascinating. I did come across Cantor's Gödel's, and Turing's work and did hear of Löwenheim-Skolem, but I've not come across *Cohen's Forcing Technique* to prove the independence of the Continuum Hypothesis and the Axiom of Choice from Zermelo-Fraenkel set theory. This alone was worth the entire experimentation exercise for me. This led me to locate Cohen's 2002 Rocky Mountain Journal paper *The discovery of forcing* ,[11], and his very readable book *Set theory and the continuum hypothesis* [12]. Reading Cohen's book led me to understanding the influence Stephen Kleene had in these matters, as well Kleene's book *Introduction to Metamathematics* [13]. So the LLMs are there quietly stirring up butterfly effects ☺.

I did find the discussion of Cantor's Diagonal Argument in Cryptanalysis very interesting , and it may actually be immensely useful in creating some quantum computing proof encryption methods, per the claim that "diagonal arguments can be used to show the existence of a key that cannot be reached by any systematic enumeration, thereby proving the security of cryptographic systems against certain types of brute-force attacks".

Obviously one can go into further detail with topics. This FLEX provided the right amount of insight for the effort.

¹⁸https://en.wikipedia.org/wiki/Berry_paradox

¹⁹https://en.wikipedia.org/wiki/Time_hierarchy_theorem

²⁰https://en.wikipedia.org/wiki/Rice%27s_theorem

4 Knowledge Hierarchies & Mapping Exercises

LLMs are intensively dependent on knowledge! While most users intuitively understand that no current LLM has trained on all available knowledge known by the human race, the actual details of what LLMs use and how they are trained is usually not explicitly understood or appreciated. I crafted this section to probe some aspects of abstract knowledge provenance and considerations. The material here reflects a guided exercises using various LLMs to discuss this considerations. The FLEX ID for this Exercise is 2024-04-08T23:17:43Z (using UTC date and time standards)²¹

⇒ Start of FLEX 2024-04-08T23:17:43Z ⇒

4.1 Knowledge Mapping

User: There are 4 levels of a) Knowledge known by an individual, b) Knowledge known by Foundational LLMs c) Knowledge known by humanity captured in texts d) Total Knowledge known by humanity at large. Tasks for you 1- Create a highly efficient and effective prompt that explores mapping and relationships between the 4 levels of knowledge, and suggested additional topics. 2 -Create the prompt so it generates further prompts to export this topic In a detailed step by step way. Use tables as necessary

To create a prompt that effectively explores and maps the relationships between the four levels of knowledge, we need to construct a framework that not only compares these levels but also generates further inquiries into their interactions and boundaries. Here's how we can structure the prompt:

Prompt for Mapping and Exploring the Relationships Between Levels of Knowledge

Objective: To analyze and map the relationships and distinctions between the four levels of knowledge: a) individual knowledge, b) knowledge possessed by Foundational Large Language Models (LLMs), c) collective human knowledge, and d) knowledge captured in texts.

Step 1: Define Each Level of Knowledge - Individual Knowledge: Insights, skills, and experiences unique to an individual. - **Foundational LLMs Knowledge:** Information and patterns learned from vast datasets by foundational LLMs like GPT-4. - **Collective Human Knowledge:** Sum of all knowledge known by humanity, including tacit and explicit knowledge. - **Knowledge Captured in Texts:** Recorded information in written or digital form.

Step 2: Establish Criteria for Comparison - Accessibility: How easily can each knowledge level be accessed by individuals or systems? - **Accuracy:** Reliability and truthfulness of the knowledge. - **Scope:** Range or breadth of the knowledge covered. - **Updatability:** How frequently and easily the knowledge can be updated.

Step 3: Create a Comparative Analysis Table Construct a table to compare the four levels of knowledge based on the above criteria.

²¹I will use UTC date and time for the FLEX ID if the topic is not yet entered in to my personal knowledge base for tracking.

Criteria	Individual Knowledge	Foundational LLMs Knowledge	Collective Human Knowledge	Knowledge Captured in Texts
Accessibility				
Accuracy				
Scope				
Updatability				

Step 4: Explore Interconnections and Flow of Knowledge - How does knowledge transfer between these levels? - What impact do advancements in one level have on the others?

Step 5: Suggest Further Topics for Exploration - The role of technology in bridging gaps between these levels. - The impact of cultural and societal factors on the flow of knowledge.

Step 6: Generate Sub-Prompts For each area identified in the previous steps, create detailed sub-prompts to explore specific aspects in depth. For example:
 - Sub-Prompt 1: Examine the impact of individual cognitive biases on the accuracy of collective human knowledge.
 - Sub-Prompt 2: Analyze the role of foundational LLMs in enhancing the accessibility of knowledge captured in texts.

Conclusion Summarize findings and propose next steps for further research or application of the mapped knowledge relationships.

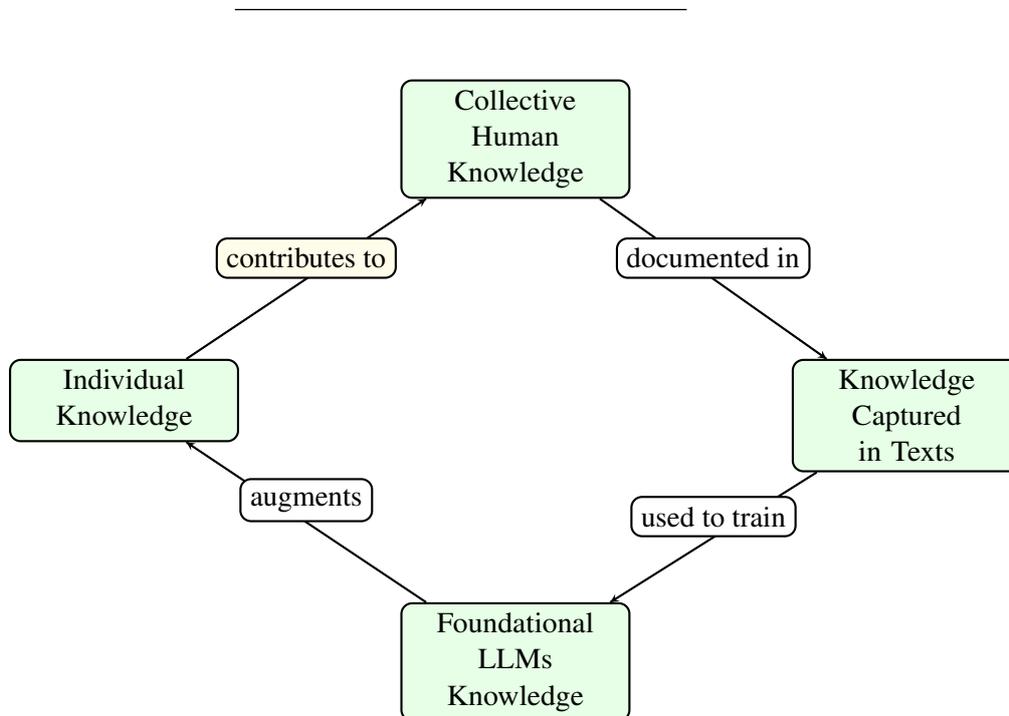


Figure 4: *Human Knowledge Availability and Access*

4.2 Beliefs and Knowledge

User: discuss in detail the beliefs, justified beliefs, justified true beliefs, stories, hallucinations and other categories of 'stuff' written in books and resident in human minds.

4.2.1 Beliefs

- **Definition:** A belief is something that an individual accepts as true. It is a cognitive state in which a person commits to a proposition or premise.
- **Characteristics:** Beliefs can be true or false, and they do not necessarily have to be based on evidence or reason. For example, a person might believe that it will rain tomorrow based on intuition rather than weather reports.
- **In Books:** Literature often presents characters' beliefs as narrative elements, which may or may not align with reality within the story.

4.2.2 Justified Beliefs

- **Definition:** A justified belief is one that a person holds on the basis of good reasons or evidence.
- **Characteristics:** Justification requires a basis for the belief, which could be empirical evidence, logical deduction, or authoritative testimony. The strength of the justification can vary.
- **In Books:** Non-fiction books, especially academic texts, strive to present justified beliefs through citations, experiments, and logical arguments.

4.2.3 Justified True Beliefs

- **Definition:** In traditional epistemology, a justified true belief is considered knowledge. It is a belief that is true and for which one has good reasons.
- **Characteristics:** The 'true' component means the belief corresponds to reality or facts. The 'justified' component means there are sound reasons that support the belief.
- **In Books:** Academic texts and authoritative references usually aim to present justified true beliefs. The process of peer review in academic publishing is meant to ensure that published knowledge is both true and justified.

4.2.4 Stories

- **Definition:** Stories are narratives or accounts of a series of events, real or imagined.
- **Characteristics:** Stories can be purely fictional, based on true events, or a mixture of both. They are integral to culture and communication and often serve to entertain, educate, or preserve history.

- **In Books:** From novels to history books, stories form the backbone of literature. Fictional stories are recognized as the product of imagination, while historical narratives strive for accuracy.

4.2.5 Hallucinations

- **Definition:** Hallucinations are perceptions in the absence of external stimuli that have qualities of real perceptions. They are vivid and clear, with the full force and impact of normal perceptions, not under voluntary control.
- **Characteristics:** They may occur in any sensory modality — visual, auditory, olfactory, gustatory, tactile, or somatic.
- **In Human Minds:** Hallucinations are associated with particular mental and neurological conditions and can also be induced by certain substances or extreme stress. They are generally not considered justified true beliefs because they do not correspond to reality.
- **In Books:** Hallucinations may be described in literature to explore characters' inner experiences or to blur the lines between reality and perception.

4.2.6 Other Categories

- **Myths and Legends:** These are traditional stories, often concerning the early history of a people or explaining some natural or social phenomenon, and typically involving supernatural beings or events.
- **Ideologies:** Systems of ideas and ideals, especially those that form the basis of economic or political theory and policy. They can strongly influence beliefs and values.
- **Propaganda:** Information, especially of a biased or misleading nature, used to promote a political cause or point of view. It may contain beliefs, but they are presented in a way to shape public opinion, not necessarily to convey truth.

In considering all these categories, it's important to note that human cognition is complex. Our minds can hold contradictory beliefs, be influenced by cognitive biases, and be swayed by compelling narratives. This makes the classification and evaluation of beliefs, knowledge, and other mental content an intricate task. Literature and non-fiction writing reflect this complexity by presenting a vast array of perspectives, stories, and purported truths.

User: User Expand in granular detail, and organize in logical subsumption hierarchies

Organizing these concepts into a logical subsumption hierarchy means we'll start with the broadest category and move down to more specific subsets. Here's a granular and detailed exploration within such a hierarchy:

4.2.7 Epistemic Content

1. Knowledge

- **Justified True Belief**

- **Empirical Knowledge:** Based on observation, experimentation, or direct sensory experience.
 - * **Scientific Knowledge:** Derives from scientific methods, including hypotheses testing and validation through experiments.
- **Rational Knowledge:** Based on reason and logical deduction.
 - * **Mathematical Knowledge:** Involves proofs and theorems derived from axioms and logical processes.
- **Authoritative Knowledge:** Based on the credibility of the source.
 - * **Historical Records:** Documented events validated by multiple sources.

2. Belief

- **Justified Belief** (Without confirmation of truth)
 - **Inferential Belief:** Derives from indirect evidence or reasoning.
 - * **Predictive Beliefs:** Expectations about future events based on current knowledge or patterns.
- **Unjustified Belief**
 - **Superstition:** Belief that is not based on reason or knowledge, in the supposed influence of some magic or supernatural force.
 - **Speculative Belief:** Hypothetical conjecture without sufficient evidence.
 - * **Conspiracy Theories:** Attempts to explain an event or situation with a hypothesis that contradicts the mainstream or established understanding.

3. False Belief

- **Inaccurate Belief:** Contradicts available empirical evidence or logical reasoning.
 - **Misconception:** An incorrect understanding or interpretation of information.

4.2.8 Narrative Content

1. Factual Narrative

- **Non-Fiction**
 - **Biography:** A detailed description of a person's life.
 - **Essay:** A short piece of writing on a particular subject, often presenting the author's personal viewpoint.
- **Historical Narrative**
 - **Chronicle:** A factual written account of historical events in the order of their occurrence.

2. Fictional Narrative

- **Fiction**

- **Novel:** A long narrative work of fiction, usually with a complex plot and character development.
- **Short Story:** A brief fictional narrative.

- **Fantasy**

- **Mythology:** A collection of myths, especially one belonging to a particular religious or cultural tradition.
- **Fairy Tale:** A children's story about magical and imaginary beings and lands.

3. Semi-Fictional Narrative

- **Historical Fiction:** Stories set in a past period that may include historical persons, but also fictional characters or events.

4.2.9 Perceptual Content

1. Veridical Perception

- **Accurate Sensory Experience:** Perceptions that align with external stimuli and are verifiable by others.

2. Non-Veridical Perception

- **Hallucination**

- **Psychologically-Induced:** Hallucinations arising from mental health conditions.
- **Substance-Induced:** Hallucinations provoked by drugs, medication, or intoxication.

- **Illusion**

- **Visual Illusion:** A misinterpretation of a visual stimulus.
- **Auditory Illusion:** A misinterpretation of auditory information, such as the McGurk effect.

4.2.10 Ideological Content

1. Political Ideology

- **Capitalism, Socialism, etc.:** A set of ethical, economic, and political beliefs about how a society should be organized.

2. Religious Ideology

- **Theism, Atheism, etc.:** Beliefs concerning the existence and nature of a deity, or the lack thereof.

3. Cultural Ideology

- **Nationalism, Feminism, etc.:** Beliefs that relate to cultural identities and societal roles.

4.2.11 Persuasive Content

1. Propaganda

- **Political Propaganda:** Information, especially of a biased nature, used to promote or publicize a particular political cause or point of view.
- **Commercial Advertising:** Information used to promote commercial products or services.

2. Rhetoric

- **Persuasive Writing:** The art of using language to persuade or influence others.
- **Debate:** A formal discussion on a particular topic in a public meeting or legislative assembly.

In this hierarchy, each category contains progressively more specific subsets, organized by their relation to truth, evidence, and presentation. This framework allows for a granular understanding of the diversity of content in human cognition and writing.

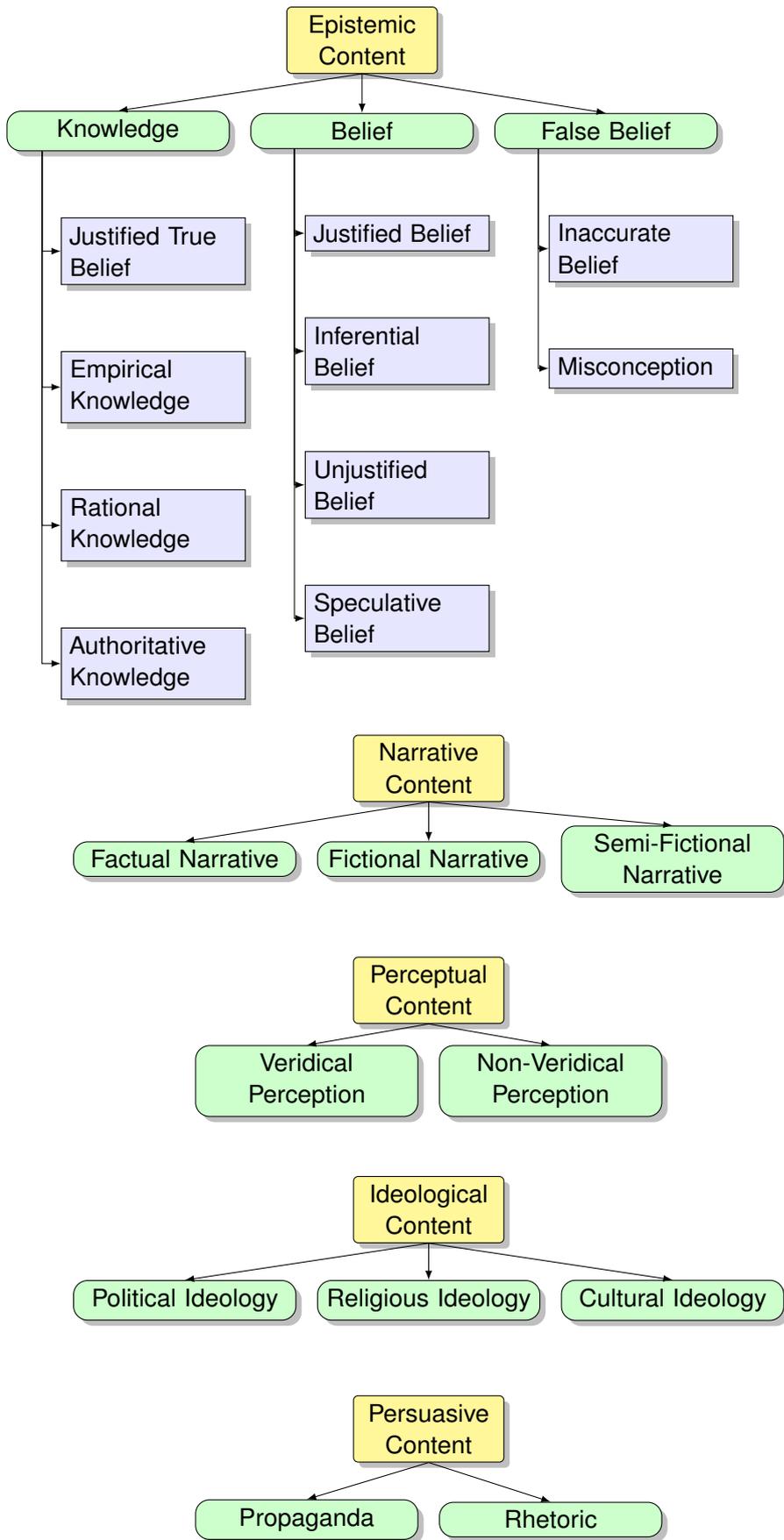


Figure 5: Knowledge Hierarchies

4.3 Discussion and observations

This FLEX was meant to get a sense of what levels of depth and granularity one can get to with a rather simple prompt. Just looking at the level of detail and breadth, it's clear that SOTA LMMs are indeed powerful, and will get increasingly so very quickly. The main challenge is in improving articulation of prompts. The real gems are to be found in discovering good questions and good topics. Each of the areas mentioned has vast amount of writings to assimilate and make sense of. Figure 4 illustrates a simplistic perspective diagramming levels and scope of Human Knowledge availability and access. Figure 5 illustrates some of the varieties of knowledge. Neither are complete, they are illustrative and indicative of the types of considerations one needs to make. Without a doubt, this area deserves much more attention and separate extensive treatment.

Clearly, the amount of Human Knowledge that exists in the brains and minds of living people, in their habits, practices, and skills is vast. Only a small amount of that is recorded at all anywhere. There are tacit skills and knowledge that people actually do not know how to express in writing. They can be shown by example, and by embedded trial and error, as in learning how to ride a bike. We can *talk* about developing a sense of balance riding a bicycle, but it's pretty impossible to teach someone how to ride a bike via text; it can only be done by riding bikes. Similarly with many physical activities whether they are dancing, playing basketball, or executing gymnastics. So there's a lot of inexpressible knowledge; one can describe activities, and check to see if certain steps are taken, but the associated deep visceral knowledge is not to be found in LLMs.

Next, though there may be a lot of knowledge written down; very little of that knowledge is actually converted into computer usable form. Some day in the future more 'stuff' will be scanned or recorded, and translated, and then be available for machine access. We're not there yet. Next come practical considerations, even if everything ever written and recorded is converted into machine accessible format, there's just too much of it. A good fraction of it is likely speculative, inaccurate, poetic, inconsistent, wishful and aspirational, and currently too expensive to process and store for LLM training. Practical considerations actually are limiting, and so developers of LLMs must choose how much to work with.

There are a variety of sources that are considered authoritative or partially so. With today's technology, one can identify the high quality scholarly, artistic, expressive, works in various media (text, images, videos, sculptures, architecture, . . .). In certain areas like medicine, there are extensive databases of all types that cover peer reviewed material, clinical trials, pharmaceutical data submitted for evaluation, and so on. The US National Library of Medicine²² (NLM) has vast collections that contain enormous amount of information in all areas relating to medicine and health care. It states that it "curates the world's richest collections of historical material from the 11th century to the most cutting-edge data of the 21st century".

²²<https://www.nlm.nih.gov>

There's *Common Crawl*²³ crawls the web every month. Its archive consists of petabytes of data collected since 2008²⁴. Open AI researchers, described using 410 billion tokens from Common Crawl in training as part of the datasets used to train GPT-3 [14]²⁵.

²³<https://commoncrawl.org>

²⁴https://en.wikipedia.org/wiki/Common_Crawl

²⁵OpenAI's paper also mentioned 19 billion tokens of WebText2, 67 billion tokens from books, and 3 billion tokens from Wikipedia; that was in 2020 current models have trained with much more data; though the details of closed models are opaque

5 Tiptoeing into AGI and SuperIntelligence

Two of the most interesting and concerning coming developments in AI are AGI (*Artificial General Intelligence*) and Superintelligence. While the idea of making very powerful AI has been around, IJ Good²⁶ articulated the idea of very powerful AI in his 1965 paper *Speculations concerning the first ultraintelligent machine*[15]. A more recent eloquent articulation of potential powers and dangers of very powerful AI were provided in Bostrom's 2014 book *Superintelligence: Paths, dangers, strategies* [16]. There are many discussions concerning risk and potential catastrophes associated with very powerful AI and AGI. These are left for another time.

I thought it would be useful to get a sense of what ChatGPT-4, Cohere Coral, and Claude Sonnet has to say regarding these matters.

⇐ Start of AGI FLEX ⇐

5.1 Experimenting with ChatCPT-4

User: Discuss this statement by I.J. Good in "Speculations Concerning the First Ultraintelligent Machine", Advances in Computers, vol. 6, 1965.

"Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultra-intelligent machine could design even better machines; there would then unquestionably be an "intelligence explosion," and the intelligence of man would be left far behind (see for example refs. [22], [34], [44]). Thus the first ultraintelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control." tasks 1) provide context 2) assess and evaluate current progress in computing and AI relative to Good's vision, 3) Provide granular discussion of AGI, 4) create a table contrasting Good's Ultraintelligent Machine, AGI, and Bostrom's SuperIntelligence

Sure, let's break down I.J. Good's statement and its implications:

1. **Context:** I.J. Good was a British mathematician who worked with Alan Turing during World War II. He was among the early thinkers to consider the potential of artificial intelligence (AI) and its implications for humanity. In his paper *Speculations Concerning the First Ultraintelligent Machine*, published in 1965, Good contemplates the concept of ultraintelligent machines, which he defines as machines capable of surpassing all intellectual activities of any human.
2. **Assessment of Current Progress:** Since Good's time, significant progress has been made in the field of AI and computing. While we have witnessed remarkable advancements in narrow AI applications, such as image recognition, natural language processing, and game playing, achieving true ultraintelligence, as Good described, remains elusive. Current AI systems excel in specific tasks but lack the general intelligence and adaptability of human intellect.

²⁶https://en.wikipedia.org/wiki/I._J._Good, Good was Turing's statistics assistant during his Enigma days; also, the idea of *The Singularity* is really derivative from Good's ultraintelligence concept.

3. **AGI (Artificial General Intelligence):** Good’s concept of ultraintelligent machines aligns closely with the notion of AGI, which refers to AI systems capable of understanding, learning, and applying knowledge across diverse domains in a manner analogous to human intelligence. Achieving AGI remains a significant challenge in AI research, as it requires not only advanced algorithms but also a deep understanding of human cognition and consciousness.
4. **Comparison with Bostrom’s Superintelligence:** Nick Bostrom, in his book *Superintelligence: Paths, Dangers, Strategies*, discusses the concept of superintelligence, which refers to intellect surpassing that of the brightest humans across all domains. While Good’s ultraintelligent machine concept focuses on surpassing human intellectual capabilities, Bostrom’s superintelligence encompasses broader implications, including the potential for AI to autonomously improve itself beyond human comprehension.

Now, let’s create a table contrasting Good’s Ultraintelligent Machine, AGI, and Bostrom’s Superintelligence:

Aspect	Ultraintelligent Machine	AGI	Superintelligence
Intellectual capabilities	Surpass all human	Generalized intelligence	Exceed human capabilities
Designing capabilities	Can design better machines	Can potentially design improved machine	Can autonomously improve beyond human limits
Current progress	Theoretical concept	Ongoing research goal	Theoretical concept
Control and governance	Requires control mechanisms to ensure safety	Concerns about control and governance	Risks associated with control and alignment

This table highlights the differences between these concepts in terms of their intellectual capabilities, design abilities, current progress, and challenges related to control and governance.



User: What explicit efforts are underway to attain AGI, who is involved and how far have they gotten?

Efforts to attain AGI involve a variety of organizations, including research institutions, universities, and technology companies. Some of the key players and initiatives in the pursuit of AGI include:

1. **OpenAI:** OpenAI is a research organization focused on advancing artificial intelligence in a way that benefits humanity. They have been

actively working on AGI research and have developed prominent models such as GPT (Generative Pre-trained Transformer) series, which have demonstrated impressive capabilities in natural language understanding and generation.

2. **DeepMind:** DeepMind, a subsidiary of Alphabet Inc. (Google's parent company), is known for its research in AI and machine learning. They are exploring various aspects of AGI through projects like AlphaGo, AlphaZero, and MuZero, which have achieved superhuman performance in games like Go, chess, and video games.
3. **IBM Research:** IBM has been conducting research in AI for decades, with a focus on developing cognitive computing systems. Their Watson AI system gained fame for defeating human champions in the quiz show Jeopardy! IBM continues to work on advancing AI technologies towards achieving AGI-like capabilities.
4. **University Research Labs:** Numerous universities around the world have dedicated research groups focusing on AGI. These include institutions such as Stanford University, Massachusetts Institute of Technology (MIT), University College London (UCL), and many others. These research groups contribute to the theoretical understanding and practical development of AGI algorithms and systems.
5. **Startups:** There are also several startups dedicated to AGI research, although they may have smaller scale compared to larger organizations. These startups often focus on innovative approaches to AGI, leveraging new algorithms, architectures, or application domains.

As for how far they have gotten, it's important to note that achieving AGI is a complex and challenging task, and there is no consensus on how close we are to realizing it. While significant progress has been made in AI research, particularly in narrow AI applications, AGI requires advancements in multiple areas such as machine learning, cognitive science, neuroscience, and more. Researchers continue to push the boundaries of AI capabilities, but achieving true AGI remains an open question with no clear timeline.



User: be significantly more explicit. Provide names of research projects and Principal Investigators

Certainly, here are some specific research projects and their principal investigators that are explicitly focused on advancing towards AGI:

1. **OpenAI:**
 - **GPT (Generative Pre-trained Transformer) series:** Led by a team of researchers including Ilya Sutskever, Sam Altman, Greg Brockman, Wojciech Zaremba, and others. Notable contributions include GPT-3, a large-scale language model with remarkable natural language understanding and generation capabilities.
 - **OpenAI AGI Safety:** This initiative focuses on ensuring the safe development and deployment of AGI systems. Key researchers involved include Dario Amodei, Paul Christiano, and Nick Cammarata.

2. DeepMind:

- **AlphaGo:** Led by Demis Hassabis, David Silver, and others, AlphaGo made headlines by defeating world champion Go player Lee Sedol. This project showcased advanced AI capabilities in strategic decision-making and planning.
- **AlphaZero:** Building upon the success of AlphaGo, AlphaZero, led by David Silver and Julian Schrittwieser, achieved superhuman performance in chess, shogi, and Go without any prior knowledge.
- **MuZero:** A project led by Julian Schrittwieser and others, MuZero is an extension of AlphaZero that learns to plan and adapt in environments without knowledge of their dynamics.

3. IBM Research:

- **Watson AI:** Although Watson is primarily known for its applications in various domains such as healthcare and finance, IBM researchers, including David Ferrucci and Eric Brown, have been involved in advancing AI technologies towards more general intelligence.
- **Project Debater:** Led by Ranit Aharonov and Noam Slonim, Project Debater is an AI system capable of debating with humans on complex topics, showcasing advancements in natural language understanding and argumentation.

4. University Research Labs:

- **Stanford University:** The Stanford AI Lab, led by Fei-Fei Li, focuses on various aspects of AI research, including computer vision, natural language processing, and robotics, with implications for AGI.
- **Massachusetts Institute of Technology (MIT):** The MIT Computer Science and Artificial Intelligence Laboratory (CSAIL), under the leadership of Daniela Rus and others, conducts research in AI and robotics, exploring fundamental principles of intelligence.
- **University College London (UCL):** The UCL Centre for Artificial Intelligence, led by David Barber and others, conducts interdisciplinary research in AI, machine learning, and cognitive science, with a focus on understanding and replicating human-like intelligence.

5. Startups:

- **Vicarious:** Founded by Dileep George and Scott Phoenix, Vicarious aims to develop AGI algorithms inspired by the human brain's structure and function.
- **Numenta:** Founded by Jeff Hawkins and Donna Dubinsky, Numenta focuses on understanding the principles of neocortical algorithms and applying them to create intelligent machines.

These projects and researchers represent just a subset of the diverse efforts aimed at advancing towards AGI, each contributing unique perspectives and

approaches to the challenge.



omment: The responses are decent, and definitely time savers for those not acquainted with the area, but they're not very deep. This may indicate the needs to design questions and prompts that are much more sophisticated. It also suggests the LLMs are tuned to users with intellectual level and language style comfortable to the 'men and women in the street'm rather than researchers and scholars. This shouldn't be too surprising, but I would hope that future adaptive LLMs can figure out users' comprehension and erudition levels.

I will repeat some of the same question for Cohere-Coral



5.2 Experimenting with Cohere Coral

In his 1965 article, British mathematician Irving John Good, who worked as a cryptologist with Alan Turing, speculated on the potential of creating an ultraintelligent machine. Good defined an ultraintelligent machine as one that could surpass all human intellectual activities and, by virtue of being able to design better machines, would lead to an intelligence explosion, leaving human intelligence far behind.

Good's vision was that this machine would be the last invention humanity would ever need to make, provided it could be controlled. He also acknowledged the potential risks, noting that it could lead to the human race becoming redundant.

5.2.1 Current Progress in Computing and AI relative to Good's Vision

Since Good's speculation, AI has become an integral part of many industries, with large companies adopting AI technologies to improve efficiency, reduce errors, and extract insights from data. AI has also been applied in medical diagnosis, economic-financial applications, robot control, law, scientific discovery, and more.

While AI has made significant progress, it has not yet reached the level of ultraintelligence as envisioned by Good. However, there have been advancements in specific areas, such as optical character recognition, speech recognition, and robotics.

In terms of AGI, there has been progress, with some speculating that current Large Language Models (LLMs) exhibit sparks of AGI. Microsoft researchers evaluated GPT-4 and concluded that it could be considered an early, yet incomplete, version of an AGI system.

5.2.2 Granular Discussion of AGI

AGI, or Artificial General Intelligence, refers to the ability of machines to mirror human capabilities across most tasks, including thinking, comprehending, learning, and solving complex problems. It aims to go beyond narrow AI, which is designed for specific tasks, and achieve a broader range of capabilities.

The journey towards AGI is often likened to climbing a mountain, with varying levels of complexity. Frameworks have been proposed to map this progress, such as DeepMind’s framework, which categorizes AGI development into distinct levels similar to the levels of autonomous driving.

5.2.3 Contrasting Good’s UltraIntelligent Machine, AGI, and Superintelligence

	UltraIntelligent Machine	AGI	Bostrom’s Superintelligence
Definition	A machine that can surpass all human intellectual activities and design even better machines, leading to an intelligence explosion.	AI that mirrors human capabilities across most tasks, learning and adapting like a human.	Any intellect that greatly exceeds human cognitive performance in virtually all domains of interest.
Current Status	Not yet realized.	In progress, with some sparks of AGI observed in LLMs.	Remains in the realm of science fiction and is not a realistic concern for experts at present.
Potential Impact	Could be the last invention humanity needs to make but also poses risks of human redundancy.	Can address scientific, social, and economic issues but also raises ethical and moral questions.	Could lead to human extinction, as feared by some.

Table 3: Cohere’s Comparisons

5.2.4 more dialog with Cohere Coral

User: What explicit efforts are underway to attain AGI, who is involved and how far have they gotten?

There are a variety of efforts underway to attain AGI, with involvement from both the private and public sectors.

Private Sector Efforts - OpenAI: Co-founded by Elon Musk, OpenAI has shifted from its original goal of being open-source and non-profit. It has developed models such as GPT-3 and ChatGPT, and is working towards creating increasingly aligned and steerable models.

-xAI: Elon Musk’s newest venture, with the goal of understanding the true nature of the universe and building a good AGI.

- DeepMind: Owned by Alphabet, DeepMind has a large team working on

alignment and was close to deploying an aligned AGI in one hypothetical scenario.

- Anthropic: CEO Dario Amodei has stated a 25% chance that their technology could end human civilization.
- Vicarious: Acquired by Intrinsic, a robotics software and AI company at Alphabet. Vicarious aims to bring about a robotic golden age by using AI to automate an increasing number of general tasks until AGI is reached.
- Microsoft: Collaborated with OpenAI in NYC to discuss balancing the risks and rewards of AI applications.
- Google: Owns DeepMind and has a strong interest in AGI.
- Meta: Chief AI Scientist Yann LeCun maintains a positive stance on AI x-risk discussion.

Public Sector Efforts - US Government: Has dominated the AI ecosystem, with surveillance capabilities to monitor who is building AGI. Has also attacked other projects with immense cyber offensive capabilities. - Chinese Government: Has copied leading lab models and decided to deploy despite alignment uncertainty.

Despite these efforts, AGI has not yet been attained, and there are differing views on how close we are to achieving it. Some predict AGI by 2025-2029, while others believe it is much further away, or may never be possible.



User: be significantly more explicit. Provide names of research projects and Principal Investigators

Here is a list of explicit efforts and key individuals involved in the pursuit of AGI:

- **OpenAI**: A non-profit AI research company co-founded by Elon Musk. OpenAI has developed models like GPT-3 and ChatGPT, and is working towards creating increasingly aligned and steerable models. Key figures include Sam Altman, Paul Christiano, Jack Clark, Holden Karnofsky, Tasha McCauley, and Nate Soares.
- **xAI**: Elon Musk's newest venture, with the goal of understanding the true nature of the universe and building a good AGI.
- **DeepMind**: Owned by Alphabet, DeepMind has a large team working on AGI and AI safety. Notable figures include Demis Hassabis, Shane Legg, and Mustafa Suleyman.
- **Anthropic**: A company working on AGI and AI safety. CEO Dario Amodei

has expressed concerns about the potential risks of their technology.

- **Vicarious:** Acquired by Intrinsic, a robotics software and AI company at Alphabet. Vicarious aims to bring about a robotic golden age by using AI to automate general tasks until AGI is reached.

- **Microsoft:** Collaborated with OpenAI and expressed interest in the development and deployment of AGI.

- **Google:** Owns DeepMind and has a strong interest in AGI.

- **Meta:** Chief AI Scientist Yann LeCun takes a positive stance on AGI and AI x-risk discussions.

- **Allen Institute for Artificial Intelligence (AI2):** Working on a project called Alexandria, which aims to bring common sense to AI by gathering and analyzing answers to common sense questions. Another project, Aristo, focuses on machine reading and reasoning in the domain of science.

- **Cerenaut:** An independent research group based in Australia that undertakes fundamental research at the intersection of AI, neuroscience, and psychology. They focus on understanding the interactions of brain regions and applying computational descriptions to benefit AI.

- **Artificial General Intelligence Society:** A non-profit organization that promotes the study and development of AGI. They host conferences, produce publications, and facilitate cooperation and communication among those interested in AGI.

- **OpenCog:** An open-source AGI project and framework led by Ben Goertzel that aims to provide a common platform for AI researchers and developers.

- **SingularityNet:** An organization described as a Webmind on blockchain, led by Ben Goertzel.

- **SoftRobot:** A Swedish AI company that utilizes machine and deep learning to create better workflow habits for businesses.

- **MIT:** Researchers at MIT are exploring various approaches to AGI, including unsupervised learning, transfer learning, common sense and causal inference, and learning optimizers.

~~~~~  
OK, lets repeat with Claude Sonnet via Poe; repeated prompts are omitted where obvious to save space. Some of the obvious caveats are also omitted.

~~~~~  
5.3 Experimenting with Claude Sonnet

To provide context, I.J. Good’s statement was made in 1965, during the early days of artificial intelligence research. At that time, the field was still in its infancy, and the idea of creating an ultraintelligent machine that could surpass human intelligence was largely speculative.

- 1) Context: Good’s statement was a thought experiment exploring the potential implications of creating an artificial intelligence system that could recursively improve itself, leading to an intelligence explosion where the AI rapidly outpaces human intelligence. This idea is now commonly referred to as the singularity concept.
- 2) Current progress in computing and AI: Since Good’s statement in 1965, there have been significant advancements in computing power, algorithms, and data availability, which have enabled the development of more sophisticated AI systems. However, we have yet to achieve the level of general artificial intelligence (AGI) or superintelligence that Good envisioned.

Current AI systems, while highly capable in specific domains, are still narrow or specialized AI systems that excel at specific tasks but lack the broad, general intelligence and self-awareness that Good described. However, the rapid progress in areas like machine learning, natural language processing, and neural networks has reignited discussions about the potential for AGI and the implications of superintelligent AI systems.

- 3) Granular discussion of AGI: Achieving AGI, an artificial intelligence system with general intelligence comparable to or exceeding human intelligence across a wide range of domains, remains an elusive goal. While there have been significant advancements in narrow AI applications, creating an AGI system that can recursively improve itself and potentially lead to an intelligence explosion as Good described is still a theoretical concept.

Current research in areas like artificial general intelligence (AGI), brain-computer interfaces, and neuromorphic computing aims to develop systems that can exhibit human-like reasoning, learning, and adaptability. However, there are still significant challenges in areas such as commonsense reasoning, general intelligence, and consciousness that need to be addressed before achieving AGI.

~~~~~  
Certainly, I can provide more explicit details and specific examples of research projects and principal investigators working in the areas of AGI and superintelligent AI systems.

|                | <b>Ultraintelligent Machine</b>                                    | <b>AGI</b>                                                                                                                               | <b>Bostrom’s Superintelligence</b>                                                                                       |
|----------------|--------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|
| Definition     | A machine that can far surpass all human intellectual activities   | An artificial intelligence system with general intelligence comparable to or exceeding human intelligence across a wide range of domains | An artificial intelligence system that greatly exceeds human cognitive capabilities in virtually all domains of interest |
| Capabilities   | Recursively improving itself, leading to an intelligence explosion | Human-like reasoning, learning, and adaptability                                                                                         | Vastly superior to human intelligence in areas like reasoning, problem-solving, creativity, and decision-making          |
| Current Status | Theoretical concept (as of 1965)                                   | Active research area, but not yet achieved                                                                                               | Theoretical concept, not yet achieved                                                                                    |
| Implications   | Potential for AI to surpass and render human intelligence obsolete | Potential for AI to match or exceed human capabilities across various domains                                                            | Potential for AI to become a transformative force, with both positive and negative impacts                               |
| Concerns       | Maintaining control over a rapidly self-improving AI system        | Ensuring AGI systems are aligned with human values and interests                                                                         | Risks of existential catastrophe or unintended consequences from superintelligent AI systems                             |

Table 4: Claude Sonnet’s Contrasting Good’s Ultraintelligent Machine, AGI, and Bostrom’s Superintelligence

- 1) Context: Good’s statement was made in the early days of AI research, before the field had made significant practical progress.
- 2) Current progress in computing and AI: While we have yet to achieve AGI or superintelligence, there have been several notable research projects and initiatives aimed at advancing these goals:
  - OpenAI’s Constitutional AI project, led by researchers like Dario Amodei and Paul Christiano, aims to develop AI systems that are corrigible and aligned with human values.
  - The Machine Intelligence Research Institute (MIRI), led by Nate Soares and Eliezer Yudkowsky, focuses on developing mathematical frameworks for understanding and mitigating potential risks associated with advanced AI systems.
  - The Center for Human-Compatible AI (CHAI) at the University of California, Berkeley, led by Stuart Russell, explores techniques for developing AI systems that are provably beneficial and aligned with human values.
- 3) Granular discussion of AGI: Several research groups and principal in-

investigators are actively working on developing AGI systems:

- The OpenCog project, led by Ben Goertzel and the SingularityNET team, aims to create a broad, general-purpose AGI system by combining various AI techniques, including neural networks, probabilistic logic, and evolutionary programming.
- The Numenta project, founded by Jeff Hawkins and Donna Dahl, is focused on developing a brain-inspired approach to AGI based on principles of hierarchical temporal memory (HTM) and cortical learning algorithms.
- The Semantic Pointer Architecture (SPA) project, led by Chris Eliasmith at the University of Waterloo, seeks to develop AGI systems by combining insights from neuroscience, computer science, and cognitive science.

|                | <b>Ultraintelligent Machine</b>                                    | <b>AGI</b>                                                                                                                               | <b>Bostrom's Superintelligence</b>                                                                                                                                             |
|----------------|--------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Definition     | A machine that can far surpass all human intellectual activities   | An artificial intelligence system with general intelligence comparable to or exceeding human intelligence across a wide range of domains | An artificial intelligence system that greatly exceeds human cognitive capabilities in virtually all domains of interest                                                       |
| Capabilities   | Recursively improving itself, leading to an intelligence explosion | Human-like reasoning, learning, and adaptability across various domains                                                                  | Vastly superior to human intelligence in areas like reasoning, problem-solving, creativity, and decision-making                                                                |
| Current Status | Theoretical concept (as of 1965)                                   | Active research area, but not yet achieved (e.g., OpenCog, Numeta, SPA)                                                                  | Theoretical concept, not yet achieved                                                                                                                                          |
| Implications   | Potential for AI to surpass and render human intelligence obsolete | Potential for AI to match or exceed human capabilities across various domains                                                            | Potential for AI to become a transformative force, with both positive and negative impacts (as explored by researchers like Nick Bostrom, Stuart Russell, and Toby Ord)        |
| Concerns       | Maintaining control over a rapidly self-improving AI system        | Ensuring AGI systems are aligned with human values and interests (e.g., work by MIRI, CHAI)                                              | Risks of existential catastrophe or unintended consequences from superintelligent AI systems (explored by researchers like Michael Cohen, Eric Drexler, and Eliezer Yudkowsky) |

Table 5: Claude Sonnet's Projects and PI answer

It's worth noting that while significant progress has been made in narrow AI applications, achieving AGI and superintelligence remains a grand challenge, with researchers exploring various approaches and techniques to address the inherent complexities and potential risks involved.



## 5.4 Deep Dive into AGI Research: Projects, Investigators, and Progress

*Below is Google's Gemini-1.0<sup>27</sup> response to the third question. I didn't repeat the whole sequence as it was repetitive. Note too that Google has subscription versions (Gemini 1.5 that presumably is better)*

Here's a more explicit look at specific AGI research efforts, the minds behind them, and the current landscape:

### Deep Learning:

- **Project: Pathway Language Model (PaLM)** (Google AI)
  - **Principal Investigators:** Anima Anandkumar, Dario Amodei, Ishaan Gulrajani, et al.
  - **Progress:** PaLM is a massive LLM trained on a dataset of text and code, achieving state-of-the-art performance in various language tasks. However, it still requires human supervision and struggles with real-world reasoning.
- **Project: Universal Transformer** (Google AI)
  - **Principal Investigators:** Ashish Vaswani, Noam Shazeer, Niki Parmar, et al.
  - **Progress:** This research focuses on developing a single, flexible neural network architecture capable of handling various tasks (like text, vision, and robotics) – a step towards more generalized intelligence.

### Reinforcement Learning:

- **Project: AlphaStar (DeepMind)**
  - **Principal Investigators:** David Silver, Julian Schrittwieser, Karen Simonyan, et al.
  - **Progress:** AlphaStar achieved superhuman performance in the complex strategy game StarCraft II. However, it requires vast amounts of training data and struggles with adapting to entirely new environments.
- **Project: OpenAI Five (OpenAI)**
  - **Principal Investigators:** Ilya Sutskever, John Schulman, Pradeep Dubey, et al.
  - **Progress:** OpenAI Five mastered the complex game Dota 2, demonstrating strategic reasoning and teamwork in a dynamic environment. However, the game's specific rules limit its generalizability.

### Large Language Models (LLMs):

- **Project: Megatron-Turing NLG (NVIDIA)**
  - **Principal Investigators:** Marjan Ghazvininejad, Angela Hocking, Mingxing Tan, et al.

---

<sup>27</sup><https://gemini.google.com/app/>

- **Progress:** Megatron-Turing NLG is one of the largest LLMs, showcasing impressive abilities in text generation and translation. Yet, factual accuracy and real-world understanding remain challenges.

- **Project: WuDao 2.0 (BAAI, China)**

- **Principal Investigators:** Huang Maci, Zhang Jiangjing, Zhou Nan, et al.
- **Progress:** WuDao 2.0 is a large Chinese LLM demonstrating strong performance in various language tasks. However, concerns exist regarding its transparency and potential biases.

### Neurocognitive Approaches:

- **Project: Human Brain Project (HBP) (European Union)**

- **Principal Investigators:** Henry Markram, Katrin Amunts, Wolf Singer, et al. (extensive international collaboration)
- **Progress:** The HBP aims to create a detailed computer model of the human brain to understand the biological basis of intelligence. While valuable for understanding the brain, translating those principles into functional AI remains a hurdle.

- **Project: Neural Information Processing Systems (NIPS) (Annual Conference)**

- **Organizers:** Vary each year (Leading researchers in AI and related fields)
- **Progress:** NIPS is a premier conference where researchers present cutting-edge work on neural networks, machine learning, and cognitive science, fostering collaboration and progress in understanding intelligence.

**It's important to remember that AGI research is a complex and collaborative effort.** While these projects highlight specific advancements, numerous other research groups and individuals are contributing to the field's progress.

The journey towards true AGI is ongoing, with significant challenges remaining. However, the dedication and innovation of researchers around the world offer a glimpse into the future possibilities of artificial intelligence.



⇌ **End of AGI FLEX** ⇌

## 5.5 AGI & SuperIntelligence Discussion and observations

We're not there yet as far Good's UltraIntelligent Machine, but we may be getting pretty close with LLMs and more sophisticated AI Models that can learn and adapt / manipulate their own code. When or whether this will occur at all is unknown but, *to me* at least, I am pretty sure it's occurring in tiny steps with the open source models that can be downloaded and experimented with by teenagers at home ☺That cat's well out of its bag, and probably pretty far away from home.

It is likely that individuals are doing pretty creative things using their home rigs, and using the large cloud based engines to fine tune and test parts of their inventions. This is likely true of large organizations as well, but whatever is done there, will not be publicly discussed.

I compared Chat-GPT-4, Coral Cohere (with web search), Claude Sonnet, Google's Gemini-1.0, and Microsoft's Copilot with the Ultraintelligent Machine, AGI, and Superintelligence prompts. The responses are stylistically pretty similar. Much of the core information is the same. GPT-4 **did not** provide as much information in some cases as the others for basic prompts, I did encounter information about people and projects I was not familiar with. Those were mostly with the smaller and newer efforts. No surprise there as the numbers of new entrants and researchers is growing significantly. I do have to note that I experimented with Google Gemini last (as it is furiously trying to catch up), and with Microsoft Bing's Copilot. By the time I got to the last two, there's really no point in noting down all the responses. I was disappointed with Copilot, although it does say it provides some GPT-4 based responses. OpenAI's subscription ChatGPT-4 was cleaner and more on topic than Copilot. Some of this may be a matter of tastes and impedance between user and machine (for those who get the analogy). It does seem as though the search engine / no-cost models are tuned to providing more simplistic answers. Probably one can elicit more sophisticated and higher scholarly and technical density, but one would have to prompt aggressively for that. Perhaps that is a good thing.

I am disappointed with GPT-4 very cautious approach. They can clearly provide much more responsive and detailed answers, but they don't. I do not know why this so.

The responses did not uncover significant new information in this area. The results are somewhat surprising in terms of the paucity of attribution to Good's seminal *Ultraintelligent Machine* vision. This is probably due the fact that his main interest was scholarly research associated with Bayesian statistics, and not in the promotion of machine intelligence or AI.

On the positive side, the LLMS pointed to some interesting AGI and AGI adjacent projects and their Principal Investigators.

We do need to remember that in all truth, we're just on the threshold of understanding AI and AGI, and that change and increased capabilities over the last one year period have dramatic, and will be even more so a year hence.

## 6 Closing Thoughts

This paper continued a few explorations motivated by during the writing of my first (QeU) paper exploring the utility of Large Language Models for one of my principal interests. I, as others, would like a better understanding of our Universe, and where it might be headed. this is an almost impossible task. Still, I enjoy reflecting about the Universe, and look for tiny piece parts of the puzzle to explore. With the emergence of the first public OpenAI, ChatGPT it appeared the right tools are coming of age. Not only are they coming of age, they're becoming more numerous and powerful at scale.

In discussing LLMs, one encounters the terms *Frontier Models*, which represent the cutting-edge of AI research, and *Foundational Models* that provide the necessary groundwork and basis for further advancements in the field. More specifically,

**Frontier Models:** - These are the cutting-edge, state-of-the-art AI models that push the boundaries of what is currently achievable in terms of performance and capabilities. - Frontier models often introduce novel techniques, architectures, or approaches that significantly advance the field. - They are typically developed by leading research institutions, tech companies, or collaborations between academia and industry. - These models are at the forefront of AI research and are often used as benchmarks to measure progress in the field.

**Foundational Models:** - Foundational models are the building blocks upon which frontier models are developed. - They represent the earlier iterations or foundational concepts of AI models that laid the groundwork for further advancements. - These models may not necessarily achieve the highest performance levels compared to frontier models, but they are crucial for establishing the fundamental principles and techniques in AI. - Foundational models serve as the basis for experimentation, refinement, and innovation in AI research. - They are often more accessible for educational purposes or for tasks where state-of-the-art performance is not necessary.

Many organizations work at summarizing and digesting progress of AI and LLMs. I am merely a user and explorer of the applications of AI and LLMS. For me, it's useful to look at some of the most authoratative and unbiased reviews, and so it is fortunate that Stanford University's AI Index Steering Committee associated with its Institute for Human-Centered AI just released its monumental *AI Index 2024 Annual Report*[17],

Summarizing AI Index's main findings:

1. **AI Performance:** AI surpasses humans in some tasks like image classification and English understanding but lags in areas like complex mathematics and planning.
2. **Industry Dominance in Research:** Industry leads in AI research, producing more notable models compared to academia, with increasing collaboration between the two.
3. **Increasing Costs:** The training costs of state-of-the-art AI models have risen significantly.
4. **Geographical Trends:** The U.S. leads in producing top AI models, followed by the EU and China.

5. **Lack of Standardization in Responsible AI Reporting:** There's a notable absence of standardized evaluation methods for responsible AI.
6. **Generative AI Investment:** Investment in generative AI has surged despite an overall decline in AI private investment.
7. **Impact on Workforce:** AI enhances productivity and work quality but requires proper oversight to avoid performance decline.
8. **AI's Role in Scientific Progress:** AI accelerates scientific discovery with applications like AlphaDev and GNoME.
9. **Increasing Regulations:** The U.S. sees a sharp rise in AI-related regulations over the past year and five years.
10. **Public Perception:** Awareness of AI's impact grows globally, accompanied by increasing nervousness, particularly in the U.S.

The report reviewed granular data and found the following staggering numbers of AI research and development projects and products for 2023, including the following:

1. Industry dominated frontier AI research and model production - In 2023, industry produced 51 notable machine learning models, academia contributed 15, and 21 notable models resulted from industry-academia collaborations in 2023.
2. In 2023, a total of 149 foundation models were released, over double the amount released in 2022.
3. Training *Frontier models* is very expensive! The AI Index estimates the training costs of OpenAI's GPT-4 at \$78 million, and \$191 million for Google's Gemini Ultra.
4. Open-source AI research project initiatives as measured by the number of AI-related projects on GitHub has grown from 845 in 2011 to over 1.8 million in 2023.
5. The number of AI publications has risen from approximately 88,000 in 2010 to more than 240,000 in 2022.

This, as I referred to in the introduction, is a contributor to the *Burden of Knowledge*. It will require us to use AI to begin to make sense of what all is occurring in AI research and development

## 6.1 Altering the Universe

Modern humans, since the advent of electromagnetism have increasingly influenced the Universe in ways that no other living organisms have. While the early days of generating electromagnetic radiation in the form of radio waves may not seem substantial, we've picked up the pace considerably. Immediate tangible influences are dramatically visible with changes of Earth and near Earth. We've definitely altered the trajectory that undisturbed evolution would have taken. We've accelerated the demise of many species and introduced many compounds that would have never come into existence absent human intervention. We've created chemical elements that would have never come into being, and had no trace before atomic and nuclear physics and high energy

accelerators came into being. We have continually been sending satellites and spacecraft into space, some venturing way beyond earth. While currently small, these will become larger. We've embarked on experiments to deflect asteroids. Who knows what the long run impact of those experiments will be. While the current experiments do not seem to affect planetary orbit stability, there are estimated to be some actual impacts of debris landing on Mars that wouldn't have otherwise occurred.

The various experiments, energetic electromagnetic emissions, and exoplanetary satellites aren't currently designed to deterministically *alter the Universe*, at some point we will likely embark on doing so purposefully.

We can definitely enumerate some of the ways we humans are altering the universe and events that might not have occurred or would occur differently without human involvement. These include:

1. **Climate Change:** Human activities, particularly the burning of fossil fuels and deforestation, have significantly altered the Earth's climate. This has led to phenomena such as global warming, melting polar ice caps, rising sea levels, and changes in weather patterns. These effects wouldn't have occurred at the same scale without human influence.
2. **Space Debris:** The proliferation of satellites and space missions has created a substantial amount of space debris in Earth's orbit. These debris pose risks to operational satellites and space missions. Without human involvement in space exploration and satellite deployment, this accumulation of debris wouldn't be as significant.
3. **Extinction Events:** Human activities, including habitat destruction, overhunting, and introduction of invasive species, have accelerated the rate of species extinction. Many species have gone extinct or are on the brink of extinction due to human activities. Without human involvement, these extinction events would likely have occurred at a much slower pace, allowing for natural adaptation and evolution.
4. **Astrobiology Contamination:** As humans venture further into space, there's a risk of contaminating other celestial bodies with Earth microorganisms. This could potentially interfere with the search for extraterrestrial life or compromise the integrity of scientific experiments conducted on other planets or moons.
5. **Light Pollution:** The widespread use of artificial lighting has led to light pollution, which not only affects astronomical observations from Earth but also disrupts ecosystems and natural habitats. Without human involvement, the night sky would be significantly darker, allowing for clearer astronomical observations and potentially different evolutionary pressures on nocturnal species.
6. **Radio Frequency Interference:** Human-made radio signals, such as those from communication devices and radar systems, create background noise that can and possibly already interferes with subtle natural biological phenomena.
7. **Nuclear Weapons Testing:** The detonation of nuclear weapons has left a lasting impact on Earth's geology and atmosphere. Fallout from nuclear tests has altered the isotopic composition of various elements in

the environment and left radioactive traces that will persist for thousands of years.

8. **Space Exploration:** Human exploration of space, including missions to the Moon, Mars, and beyond, has left a lasting mark on the universe. From the footprints of astronauts on the lunar surface to the rovers exploring the Martian terrain, these are events that wouldn't have occurred without human involvement and have forever changed our perspective on the cosmos.

Ideas about planetary defense for protecting Earth from asteroids and comets have been discussed in detail for sometime now [18]. Fenucci and Carbognani [19] recently reported on their analysis of The DidymosDimorphos binary asteroid NASA *Double Asteroid Redirection Test* (DART) that experimented with the kinetic impactor technique for asteroid deflection. The DART spacecraft impacted Dimorphos on 2022 September 26, 2022.

Their report stated that the Hubble Space Telescope (HST) found 37 boulders, not previously observed, and assumed to be ejected during DART's collision with Dimorphos. These boulders have with estimated diameters ranging from 4 to 7 meters. It appears the boulders are moving away from the DidymosDimorphos system. According to their analysis:

By using numerical simulations, we propagated the orbits of a swarm of boulders for 20 kyr in the future, and found that they cross the orbit of Mars.

...

Given the rarefaction of the Martian atmosphere, we expect the boulders to arrive intact on the ground and excavate a small impact crater.

While, the calculations do not definitively demonstrate the boulders will hit Mars, o me, this very much indicates the ease of, simplicity, and inevitability human material impact on the Universe. the DART experiment was essentially a very minimal experiment in nudging an off Earth asteroid, future experiments will be more substantial. Yes we have arrived at the point of demonstrating material power over (right now minuscule ) parts of the Universe.

Pierre-Simon Laplace, Henri Poincaré: Poincaré, Carl Friedrich Gauss and Joseph-Louis Lagrange are among the luminaries who investigated celestial mechanics an stability in the solar system. Currently the orbits of the planets, moons, and other bodies our solar system exhibit remarkable regularity and predictability, and things seem pretty stable. With the power to nudge and later move asteroids, we definitely will have more substantial power over the Universe. With AI, and AGI, we might even be tempted to look for more beneficial arrangements that would minimize human life exterminating asteroids. Something the dinosaurs couldn't do ☺

## 6.2 It's about *the Understanding*

Returning to the starting point, this Focused LLM Experiment is about the progressions and emergence from Brain to Mind to Understanding and ultimately to Power over the Universe. The particulars discussed were intended to provide a little bit more illumination on *my experiences* with AI, and in

particular Large Language Models. I do not claim major insights ... I am only starting to play with today's equivalent of *Mendel's Peas* ☺

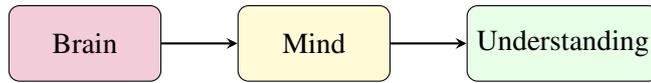


Figure 6: *from Brain to Understanding*

Figure 6, is a simple diagram to draw, but possibly one of the most challenging to explain. We can speculate about the beginnings and end of the Universe, but we can't really 'touch' them. We're too far removed from either. Brain, Mind, and Understanding are 'things' that we deal with everyday.

There's another section to this paper, section 7, that provides background notes for those interested. The substantive topics I wanted to cover here were experimenting with LLMs in arenas touching Knowledge and Understanding. I would like to close with the words of one of the world's most articulate logicians, and a friend of Bridgman, *Willard Van Orman Quine* <sup>28</sup>, liked to discuss the purpose of science. In *Pursuit of truth* [20], he noted his views of the relationship between *Understanding* to science:

Note that prediction is the main purpose of science. *One major purpose is understanding. Another is control and modification of the environment. Prediction can be a purpose too*, but my present point is that it is the **test of a theory**, whatever the purpose.

he also reminds that

combination of words often involves the imposition of additional restrictions by the mere form of the combination. The combination may, for example, take the form of a statement, in which case it is implied that a truth value attaches to the statement. Possibly there is the further implication that there is some method of verifying the truth of the statement. Or the combination may take the form of a question.

### 6.3 Final Thoughts

This paper got to be a lot longer than I intended. I was initially thinking of a much shorter paper, but it seemed like almost every day or two days, new LLM capabilities were brought to life. I also found myself digging deeper into various aspects as the the response tokens were streaming on my screen. There are many lessons to be learned here:

- Each person will respond differently to LLM outputs based on their existing knowledge and skill sets.
- One should be open to an interactive approach. Information you didn't expect will lead you to questions you haven't planned on asking, and reveal unexpected threads and directions.
- One should have some structure and methodology to work from, and then learn from the experience and update future structure methodology.

---

<sup>28</sup>[https://en.wikipedia.org/wiki/Willard\\_Van\\_Orman\\_Quine](https://en.wikipedia.org/wiki/Willard_Van_Orman_Quine)

- Different LLMs provide a certain amount of overlapping information, and a fair amount of unique information. This reflects different LLMs data sets and training and aligning methodologies. For serious research, one should consider using three totally different LLMs and then figure a way to merge and reconcile results. This is an obvious area for a talented group to create a MetaLLM. While Mistral's *Mixture of Experts* approach aims to do that, I believe that is not sufficient.
- The field is evolving and will continue to evolve. There's a lot not yet known about the Universe. Chaitin argues that there are an infinity of mathematical truths, and that we sometimes randomly stumble into them. We don't yet know which will be important truths and which will be irrelevant to humans (and the Universe).

We should keep in mind that human cognitive abilities are limited. These limitations do and will hinder quantum leaps of knowledge discovery and understanding. We need more intense development of AI to improve our understanding. We definitely could use major improvements of AI to accelerate research in mathematics and logic. We also need continuing improvements in making us, humans, better able to communicate with AI and AGI. This will be of increasing, if not critical, importance..

## 7 Notes

These end notes are provided for convenience in case you want some backstory or origins story on some topics.

### 7.1 United Federation of Galaxies' Pizza Delivery Scooter Enterprise

I really enjoyed the *Star Trek* TV series and later, the movies. I adapted the Federation Starship Enterprise idea into a space scooter labeled *United Federation of Galaxies' Pizza Delivery Scooter Enterprise*. I crafted the following narrative, and am occasionally using references to this narrative in this and future papers.

EXISTENCE: THE ULTIMATE FRONTIER. THESE ARE THE MEMOIRS OF THE UNITED FEDERATION OF GALAXIES' PIZZA DELIVERY SCOOTER *Enterprise*. ITS CONTINUING MISSIONS: TO EXPLORE CONCEPTS, AREAS, AND PLACES WHERE MANY HAVE TRODDEN BEFORE AND TO EXTRACT COHERENCE AND NEW INSIGHTS THEREOF, AND TO FIND NEW POCKETS OF REALITY, POSSIBILITIES, AND CONSTRUCTS<sup>29</sup>!



For the curious, *Eternity's Pizza* is a fictional Pizza Joint about 42 steps from *Milliways*, also known as Douglas Adams' Restaurant at the End of the Universe. It hosts, and possibly reincarnates, the convergence of Adams' and

<sup>29</sup>the choice of the *PDS Enterprise* as the transport vehicle is in keeping with the spirit of <https://eternityspizza.substack.com>

Kurt Vonnegut's visions. It houses devices that can access *Chrono-Synclastic Infundibula (CSI)*. It exists at *the end of time (TEOT)* for the ascended savants, time travelers, interdimensional beings, and future entities we, limited humans of the 21st century can hardly imagine. Located, as it were and always will be at TEOT, it fosters a special camaraderie of cultures, concepts, and beings that transcends temporal and tribal boundaries.

CSI allow converged cross temporal nebular banquets to be held with beings from different time periods and galaxies. Yup, prehistoric appetizers and delicacies from a long-lost alien civilization and argue with co-existing minds across multiple times and places all at once. The Pizza Delivery Scooter *Enterprise*, uses the *Chrono-Synclastic Infundibular Drive (CSID)* to facilitates information and material access to all times and places. The rest of the story comes later ☺.

## 7.2 Large Language Models used in this paper

I used a number of large language models (LLMs) in writing this and some other papers. The quality and output vary with time. Usually they get better, but sometimes they regress. I try to keep tab of the date of use. It's not always clear what version of what model I use. Where possible I also keep note of the URL associated with output (if and when the providers provide a unique URL; often they do.) The following is a table of LLMs I use, some much more often than other. I currently use the paid version of *OpenAI's ChatGPT-4*. One should note that there's not guarantee that if a model is available now that it won't be retired or 'deprecated' later. The models I use in my papers are mostly one or another of the models listed in table 6. The models are listed in pseudo-random order.

Hugging Face<sup>30</sup> (HF) provides information on almost 600,000 models, some of which can be accessed for use. Many of them are research models not in wide use. HF also provides access to datasets and other files, Depending on use cases and individual model policies, models and datasets, including training datasets, can be downloaded for experimentation and use. One should note though, that these are usually meant for folks who know how to code and modify LLMs. Again many are used for research purposes and studies.

---

<sup>30</sup><https://huggingface.co/models>

| LLM Name     | URL                                                                                                                                    |
|--------------|----------------------------------------------------------------------------------------------------------------------------------------|
| ChatGPT      | <a href="https://chat.openai.com">https://chat.openai.com</a> - currently running GPT-4                                                |
| Meta Llama-3 | <a href="https://meta.ai">https://meta.ai</a> - Various LLMs; Llama3-70B is the shining examle                                         |
| Coral        | <a href="https://coral.cohere.com">https://coral.cohere.com</a> - Cohere Coral Chat / Search / documents                               |
| Groq         | <a href="https://groq.com">https://groq.com</a> - Various LLMs including Gemma-7B-LT, Llama3-70B-8192, Mixtral-8x7B-32768, and more    |
| Perplexity   | <a href="https://labs.perplexity.ai">https://labs.perplexity.ai</a> - choice of 14 models, DBRX                                        |
| phind        | <a href="https://www.phind.com/search?home=true">https://www.phind.com/search?home=true</a> , runs phind-34B, 70-B, Claude Opus, GPT-4 |
| Poe          | <a href="https://poe.com">https://poe.com</a> - all Claude versions, and many other LLMs                                               |
| Copilot      | Microsoft Bing's family of models <a href="https://copilot.microsoft.com/">https://copilot.microsoft.com/</a>                          |
| Gemini       | Google's family of models <a href="https://gemini.google.com/app">https://gemini.google.com/app</a>                                    |

Table 6: Some of the Principal Large Language Models in Current Use

### 7.3 Using Tikz for Diagrams and TableMaking

I'm including this short discussion and some illustrations of using LaTeX Tikz code for those not familiar with it. For some reason, I didn't utilize LaTeX until a few years ago. Since I started composing work in LaTeX, I've grown more and more fond of it, and I appreciate all the hard work the open-source community has done with it.

I include a few examples of making some diagrams and tables here in case you want a glimpse before diving in, ☺

I will list the code that I use and some diagrams generated with it.

```
\begin{center}
  \tikzset{
    event/.style={rectangle, rounded corners, draw, align=center, minimum height=2em},
    arrow/.style={->, >=stealth', shorten >=1pt, thick},
    dashedarrow/.style={->, >=stealth', shorten >=1pt, thick, dashed},
  }
  \begin{tikzpicture}[node distance=2cm, auto]
    % Nodes
    \node[event] (A) {A};
```

```

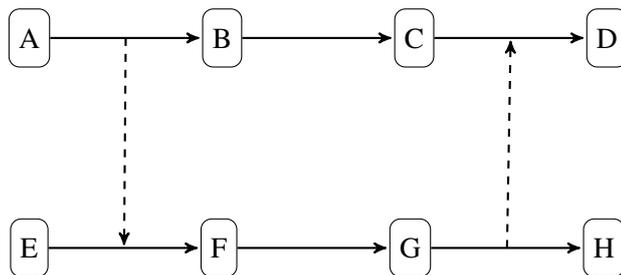
\node[event, right=of A] (B) {B};
\node[event, right=of B] (C) {C};
\node[event, right=of C] (D) {D};
\node[event, below=of A] (E) {E};
\node[event, right=of E] (F) {F};
\node[event, right=of F] (G) {G};
\node[event, right=of G] (H) {H};

% Arrows
\draw[arrow] (A) - (B);
\draw[arrow] (B) - (C);
\draw[arrow] (C) - (D);
\draw[arrow] (E) - (F);
\draw[arrow] (F) - (G);
\draw[arrow] (G) - (H);

% Dashed arrows
\draw[dashedarrow] ($(A)!0.5!(B)$) - ($(E)!0.5!(F)$);
\draw[dashedarrow] ($(G)!0.5!(H)$) - ($(C)!0.5!(D)$);
\end{tikzpicture}
\end{center}

```

creates the following diagram



The following code creates a flow diagram:

```
\begin{center}
\begin{figure}[hb]
\begin{adjustbox}{valign=c}
\begin{tikzpicture}
[
module/.style={
draw,
rounded corners,
minimum width=2cm,
minimum height=1cm,
align=center
},
arrow/.style={
thick,
->,
>=stealth
}
]
% Define the modules
\node[module, fill=purple!20] (one) {Bob};
\node[module, fill=yellow!20, right=of one] (two) {John};
\node[module, fill=green!10, right=of two] (three) {Mike};
\node[module, fill=yellow!60, right=of three] (four) {Charlie};

% Connect the modules with arrows
\draw[arrow] (one) - (two);
\draw[arrow] (two) - (three);
\draw[arrow] (three) - (four);
\end{tikzpicture}
\end{adjustbox}
\caption{\textit{Bob to Charlie}}
\label{fig:bobtocharlie}
\end{figure}
\end{center}
```

creates this diagram:

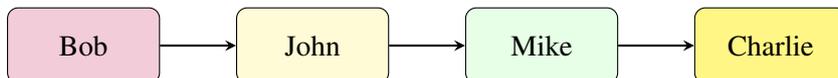


Figure 7: *Bob to Charlie*

## Tables Tables Everywhere

This code is an example of *TableMaking* ☺

```

\begin{table}[H]
\centering
\caption{Your caption.}
\label{demo-table}
\begin{tabular}{llrrrrr}
\toprule
{} & {} & {\textbf{\#d}} & {\textbf{\#i}} & {\textbf{\#l}} & {\textbf{\#s}} & {\textbf{\#o}} \\
{\textbf{Dataset}} & {\textbf{Model}} & {} & {} & {} & {} & {} \\
\midrule
\multirow[c]{5}{*}{\textit{D (simple)}} & b & 1,026 & 692 & \itshape 31 & 284 & 1,007 \\
& f & 1,366 & 398 & 49 & 238 & 685 \\
& f+d & 1,372 & 415 & 46 & 215 & 676 \\
& f+d+pt & 1,372 & 415 & 46 & 215 & 676 \\
& f+d+tg & \bfseries 1,732 & \itshape 8 & 82 & 262 & \itshape 352 \\
\multirow[c]{5}{*}{\textit{D (complex)}} & b & 527 & 266 & 228 & 589 & 1,083 \\
& f & 759 & 153 & 270 & 470 & 893 \\
& f+d & 749 & 177 & 233 & 456 & 866 \\
& f+d+pt & 749 & 177 & 233 & 456 & 866 \\
& f+d+tg & \bfseries 969 & \itshape 33 & \itshape 38 & \itshape 380 & \itshape 451 \\
\bottomrule
\end{tabular}
\end{table}

```

Table 7: Your caption.

|                    |        | <b>#d</b>    | <b>#i</b> | <b>#l</b> | <b>#s</b> | <b>#o</b> |
|--------------------|--------|--------------|-----------|-----------|-----------|-----------|
| <i>D (simple)</i>  | b      | 1,026        | 692       | 31        | 284       | 1,007     |
|                    | f      | 1,366        | 398       | 49        | 238       | 685       |
|                    | f+d    | 1,372        | 415       | 46        | 215       | 676       |
|                    | f+d+pt | 1,372        | 415       | 46        | 215       | 676       |
|                    | f+d+tg | <b>1,732</b> | 8         | 82        | 262       | 352       |
| <i>D (complex)</i> | b      | 527          | 266       | 228       | 589       | 1,083     |
|                    | f      | 759          | 153       | 270       | 470       | 893       |
|                    | f+d    | 749          | 177       | 233       | 456       | 866       |
|                    | f+d+pt | 749          | 177       | 233       | 456       | 866       |
|                    | f+d+tg | <b>969</b>   | 33        | 38        | 380       | 451       |

**CTAN** - The Comprehensive TEX Archive Network <sup>31</sup> provides more information than one can digest in a lifetime. It provides description of  $\text{\LaTeX}$  packages usages and manuals. It states "CTAN has currently 6582 packages. 2981 contributors have contributed to it. Most of the packages are free and can be downloaded and used immediately." I learned a lot there, on more things and possibilities than I had imagined. You might it interesting and useful as well.

<sup>31</sup><https://www.ctan.org>

## 8 Dedication, Acknowledgments, and References

**Dedication** The work is dedicated to the memory of my son Brélan Moritz who departed the Earthly realm to an unknown realm in April 2015. I miss him greatly. Brélan was curious about life from a young age. After seeing the movie *Jurassic Park* as a young kid, he set out to learn all he could about DNA. His goal was to be able to resurrect dinosaurs. He became a proficient molecular biologist who could and did engineer various strains of *E. Coli*.

**Acknowledgments Declarations** The author uses contemporary a variety Large Language Models (LLMs) in this work. The names and web URLs for those are listed in section 7.2. The ones most frequently used are OpenAI’s ChatGpt-4, Claude Sonnet and Opus, Mixtral-8x7B, and Cohere. The work itself is drafted using  $\LaTeX$ , TeXstudio 4.7.3 using Qt Version 6.6.0. Markdown formatted results generated by LLMs are converted to LaTeX using the Pandoc utility.  $\LaTeX$ graphics are coded using the *TikZ* package. The manuscript was formatted with Tufte adjacent style sheets.

While no financial support was provided for research and publication of this work, the author gratefully acknowledges the contributions of all the scholars and technologies that make finding, interpreting, and using an immense amount of knowledge and data both easy and enjoyable.

**Conflict of Interest** The author declares there’s no conflict of interest associated with this work.

**Typos & Errors** There are likely some, hopefully not too many, typos in the paper ☺. There may be some errors too. I am not perfect. My apologies in advance. Let me know if you find something really really off.

## References

- [1] Pamela McCorduck. *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. CRC Press, 2004.
- [2] Charles Darwin. *The voyage of the Beagle*. Floating Press, 1839. URL: <https://bit.ly/3xeJzZX>.
- [3] Elan Moritz. *Chatting with Chat (GPT-4): Quid est Understanding?* In: (2024). URL: <https://philpapers.org/go.pl?aid=MORCWC-2>.
- [4] Percy Williams Bridgman. *Quo vadis*. In: *Daedalus* 87.1 (1958), pp. 85–93. URL: <http://www.jstor.org/stable/20026428>.
- [5] Percy W Bridgman. *Remarks on the present state of operationalism*. In: *The Scientific Monthly* 79.4 (1954), pp. 224–226.
- [6] Hartley Rogers Jr. *Theory of recursive functions and effective computability*. MIT press, 1987.
- [7] Kurt Gödel. *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*. In: *Monatshefte für mathematik und physik* 38 (1931), pp. 173–198.
- [8] Georg Cantor. *Ueber eine elementare Frage der Mannigfaltigkeitslehre*. In: *Jahresbericht der Deutschen Mathematiker-Vereinigung* 1 (1890), pp. 72–78.
- [9] Alan Mathison Turing. *On computable numbers, with an application to the Entscheidungsproblem*. In: *J. of Math* 58.345-363 (1936), p. 5.
- [10] George S Boolos, John P Burgess, and Richard C Jeffrey. *Computability and logic*. Cambridge university press, 2002.
- [11] Paul J Cohen. *The discovery of forcing*. In: *The Rocky Mountain journal of mathematics* 32.4 (2002), pp. 1071–1100.
- [12] Paul J Cohen. *Set theory and the continuum hypothesis*. Addison-Wesley, 1966.
- [13] Stephen C. Kleene. *Introduction to Metamathematics*. Noordhoff, Groningen; North-Holland Publishing Co., Amsterdam, 1952.
- [14] Tom B. Brown et al. *Language Models are Few-Shot Learners*. 2020. arXiv: 2005.14165 [cs.CL].
- [15] Irving John Good. *Speculations concerning the first ultraintelligent machine*. In: *Advances in computers* 6 (1966), pp. 31–88.
- [16] Nick Bostrom. *Superintelligence: Paths, dangers, strategies*. Oxford University Press, 2014.
- [17] Nestor Maslej et al. *The AI Index 2024 Annual Report*. Stanford, CA: AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, 2024.
- [18] Nikola Schmidt. *Planetary defense: Global collaboration for defending earth from asteroids and comets*. Springer, 2018.
- [19] M Fenucci and A Carbognani. *Long-term orbital evolution of dimorphos boulders and implications on the origin of meteorites*. In: *Monthly Notices of the Royal Astronomical Society* (2024), stae464.
- [20] Willard Van Orman Quine. *Pursuit of truth*. Harvard University Press, 1990.

**FIN**

⇐