

# Chatting with Chat(GPT-4): Quid est Understanding?

Alt Title: On the path to AGI, Loqui cum Chat (GPT-4): Quid est Intellectus?

Elan Moritz

Philadelphia, Pa.

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

---

## ARTICLE INFO


*Keywords:* AI, AGI, Understanding Artificial Intelligence, Superintelligence, ChatGPT, GPT-4, LLM, Large Language Models, Brain, Mind, Consciousness, Life, Emergence, MetaSystems Transitions, Prompts, Metaprompts, UltraIntelligent Machines, Universe. OODA Loop, UDA Arc.

2024-03-01

## ABSTRACT

What is Understanding? This is the first of a series of Chats with OpenAI's ChatGPT (Chat). The main goal is to obtain Chat's response to a series of questions about the concept of 'understanding'. The approach is a conversational approach where the author (labeled as user) asks (prompts) Chat, obtains a response, and then uses the response to formulate followup questions. David Deutsch's assertion of the primality of the process / capability of understanding is used as the starting point. Understanding is posited as a major step in the Big Picture description of emergence in the Universe. Understanding follows the emergence of Brains and Minds, culminating in Power Over the Universe. 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. This approach 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 designers of Advanced General Intelligence (AGI). Other findings include the need for creating new metrics and ways of discussing intelligence, and various adjacent areas and subprocesses associated with intelligence. Discussions of future directions is also included.

---

©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. 

---

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

**Contents**

- 0.1 Intelligence, where and what art thou? . . . . . 4
- 0.2 On the Understanding ... . . . . . 5
- 0.3 David Deutsch? . . . . . 5
- 0.4 Defining Understanding . . . . . 6
  
- 1 The Big Picture . . . . . 11**
- 1.1 Brains to Minds to Understanding . . . . . 13
- 1.2 From Understanding to Power Over the Universe . . . . . 14
- 1.3 But can we really have Power Over the Universe? . . . . . 15
  
- 2 Chatting with Chat . . . . . 18**
- 2.1 Starting Line: Deutsch’s remarks on Understanding . . . . . 20
- 2.2 On Artificial Intelligence and Understanding . . . . . 23
- 2.3 AI SOTA Limitations . . . . . 25
- 2.4 Evidence of Uderstanding by SOTA AI . . . . . 26
- 2.5 Objective Evaluation Benchmarks . . . . . 28
- 2.6 Artificial General Intelligence (AGI) and Understanding . . . . . 34
- 2.7 Understanding and Explanation: What’s the difference? . . . . . 36
- 2.8 Desperately Seeking Understanding . . . . . 37
- 2.9 Looking for books with understanding in their title . . . . . 37
- 2.10 Discussing of Locke’s *An Essay Concerning Human Under-*  
*standing* . . . . . 43
  
- 3 Brass Tacks: Getting down to business . . . . . 45**
- 3.1 Round One . . . . . 45
- 3.2 Round Two . . . . . 49
  
- 4 Assessing Chat’s Understanding of Understanding . . . . . 51**
  
- 5 Closing Thoughts . . . . . 52**
  
- 6 Dedication and Acknowledgments . . . . . 56**

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>2</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 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 ☺.

---

<sup>2</sup>the choice of the *PDS Enterprise* as the transport vehicle is in keeping with the spirit of <https://eternityspizza.substack.com>

“AI began with an ancient wish to forge the gods.”  
Pamela McCorduck, *Machines Who Think*<sup>3</sup>

## 0.1 Intelligence, where and what art thou?

### Desperately Seeking ~~Susan~~ Intelligence

In the 1985 American comedy-drama *Desperately Seeking Susan*<sup>4</sup>, Roberta becomes obsessed with a free-spirited woman named Susan, played by Madonna. Roberta immerses herself in Susan’s world after suffering from amnesia. The film explores various themes including that of seeking something driven by deep curiosity and the quest for something beyond the ordinary and mundane.

For over fifty years I’ve been driven by deep curiosity about intelligence, the brain, and the universe. Many of my questions are old questions, some are new, and some are perhaps speculatively original. The questions are of the kind philosophers and scientists have asked for millennia. We are fortunate now to have incredibly powerful and more precise tools, and astounding capability of calculation and computing. Those enable much more accurate and deeper explorations of physical reality and conceptual possibilities. The almost daily incredible advances are motivating me to dig deeper and reflect more intensely on where we are, what we know, and where we might/could/should be heading.

By habit, I tend to write to a scientific technically oriented reader. This takes way too long ... and so I decided to just capture thoughts, ideas, and experiments in a ‘lite’, informal, not-too technical way. Yes, you’ll encounter technical jargon and maybe some technical notation, but don’t let that stop you from thinking and reflecting ... do whatever works for you. Maybe something will pop and grab your interest and then motivate you to dig deeper and maybe so something totally new with it.

In the beginning ☺, I was going to lay everything out in some organized way, but SO much is happening, that I decided the best way is just to write, and then at some later point pull a few pieces together into something more polished. This ain’t it.

Basically, the grand theme is to share thoughts and ideas that I come across or pop up in my head as part of my adventures and reflections about intelligence (natural and artificial), thinking, and the mind. Various aspects of physical reality, neuroscience, biology, language, logic, civilization, physical sciences, engineering, human development and habits, societies, cultures, and more show up along the way. I sincerely hope you find something useful and perhaps inspiring here.

This particular paper is about this thing called *understanding*.

---

<sup>3</sup>McCorduck, P. (2009). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. AK Peters/CRC Press.

<sup>4</sup>directed by Susan Seidelman and starring Rosanna Arquette as Roberta Glass

## 0.2 On the Understanding . . .

This particular paper is about this thing called *understanding*. It was motivated by my recent encounter with David Deutsch's post on X<sup>5</sup>

"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.**"

So what is *understanding*? It occurred to me that it would be fun to summon the power of AI, in the form of OpenAI's GPT-4 (henceforth *Chat*)<sup>6</sup> and ask it questions about this *understanding* thing. My approach is to record my prompts and the replies. I do not vouch for the accuracy of Chat's responses. I will occasionally comment about some of the responses, and will provide some observations afterwards. As with all AI tools, one must remember they are not perfect or necessarily correct. As one has found with most, if not all Large Language Models (LLMs), and GPTs, there's a fair amount of hallucinations. To be perfectly honest, humans tend to make up things as well.

## 0.3 David Deutsch?

Why pay attention to David Deutsch? It's likely that you never heard of Deutsch<sup>7</sup>. He's kinda smart, but most importantly he asks really really good questions, and then goes and discusses those very coherently and compellingly. He knows a lot of physics, math, and logic. He also knows many other things, but mostly importantly, in all his work, he reasons lucidly and very rationally. His two very insightful and accessible books, *The Fabric of Reality*<sup>8</sup> and *The Beginning of Infinity*<sup>9</sup> discuss many of his ideas.

He explores fascinating areas in his research, for example, in one of his landmark paper on quantum computing he writes<sup>10</sup> "A class of problems is described which can be solved more efficiently by quantum computation than by any classical or stochastic method. The quantum computation solves the problem with certainty in exponentially less time than any classical deterministic computation." He also speculates convincingly about the structure of the multiverse? Sounds like he's thought not only about one universe, but many of them simultaneously<sup>11</sup>. All good and necessary to open one's imagination aperture, as well as getting prepared for possibilities not yet encountered.

Along the way he was awarded the 2023 Breakthrough Prize in Fundamental Physics (with Charles Bennett, Gilles Brassard and Peter Shor) for "founda-

---

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

<sup>6</sup>GPT stands for Generative Pre-trained Transformer

<sup>7</sup>[https://en.wikipedia.org/wiki/David\\_Deutsch](https://en.wikipedia.org/wiki/David_Deutsch)

<sup>8</sup>Deutsch, D. (1998). *The fabric of reality*. Penguin UK.

<sup>9</sup>Deutsch, D. (2011). *The beginning of infinity: Explanations that transform the world*. Penguin UK.

<sup>10</sup>Deutsch, D., & Jozsa, R. (1992). Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 439(1907), 553-558.

<sup>11</sup>Deutsch, D. (2002). The structure of the multiverse. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 458(2028), 2911-2923.

tional work in the field of quantum information”. Also awarded the Isaac Newton Medal and Prize of the Institute of Physics for “founding the discipline named quantum computation and establishing quantum computation’s fundamental idea, now known as the ‘qubit’ or quantum bit” (in 2021). He was also recognized in 2008 with being elected for Fellowship of the Royal Society 2008 with a citation that reads:

“David Deutsch laid the foundations of the quantum theory of computation, and has subsequently made or participated in many of the most important advances in the field, including the discovery of the first quantum algorithms, the theory of quantum logic gates and quantum computational networks, the first quantum error-correction scheme, and several fundamental quantum universality results. He has set the agenda for worldwide research efforts in this new, interdisciplinary field, made progress in understanding its philosophical implications (via a variant of the many- universes interpretation) and made it comprehensible to the general public, notably in his book *The Fabric of Reality*.”

Of course there are many other enlightened thinkers, he’s one of the best contemporary ones, and re-reading his ideas and questions about understanding is in fact refreshing and motivating. I will also include other points of view. I will, however, start my chat with ChatGPT with Deutsch’s take on understanding.

## 0.4 Defining Understanding

My use of the term *understanding* will be doing some pretty heavy lifting in this paper. I used it so far without definition. We all kinda understand what *understanding* means, or do we?

To ground the discussion lets start adding some definitions<sup>12</sup> to get a better sense how understanding is used in some established references.

Merriam- Webster (M-W) offers the following<sup>13</sup>: **understanding**, noun  
Synonyms of understanding

- 1.a: mental grasp : COMPREHENSION
- 2.a: the power of comprehending especially : the capacity to apprehend general relations of particulars
- 2.b: the power to make experience intelligible by applying concepts and categories
- 3.a: a mutual agreement not formally entered into but in some degree binding on each side
- 3.b: friendly or harmonious relationship
- 3.c: an agreement of opinion or feeling : adjustment of differences
- 4: EXPLANATION, INTERPRETATION
- 5:: SYMPATHY sense 3a

---

<sup>12</sup>This, and really all of this paper, ‘lives’ in the American English linguistic and cultural space. One must recognize that what’s expressed in American English will not translate accurately across all other cultures and languages. The idea of understanding of course is necessary in all normal human cognitive an social functions

<sup>13</sup>“Understanding.” Merriam-Webster.com Dictionary, Merriam-Webster, <https://www.merriam-webster.com/dictionary/understanding>, Accessed 29 Feb. 2024

*understanding*, adjective

1: endowed with understanding : TOLERANT, SYMPATHETIC an understanding supervisor

2: archaic : KNOWING, INTELLIGENT

**understandingly**, adverb

**Synonyms:**

**Nouns:** accord, agreement, bargain, compact, contract, convention, covenant, deal, disposition, pact, settlement.

*Adjectives:* commiserative, compassionate, empathetic, empathic, humane, sympathetic.

**Examples of understanding in a Sentence:**

Noun:

- He has a thorough understanding of the subject.
- She has a basic understanding of the process.
- I hope to get a clearer understanding of the issues involved.

M-W has more information and a rather extensive list of examples at various levels of use (including kids' level).

There's of course the preeminent scholarly *Oxford English Dictionary* (OED) that many regard as the accepted English Language authority in these matters. OED's main definitions of *understanding* are<sup>14</sup>:

**understanding**, in *understand*, v. To comprehend; to apprehend the meaning or import of; to grasp the idea of.

**understanding**, n. (Without article.) Power or ability to understand; intellect, intelligence.

**understanding**, adj. Of a person, etc.: displaying sympathetic tolerance; of a forgiving nature or temperament.

Of course, the OED has a lot more. If one really really wants to get into definitions and usages ... this is what they listed as available results for "understanding" in:

- Quotation work title (1,822)
- Quotation text (1,762)
- Meanings (27)
- Definitions (641)
- Etymologies (72)
- Historical Thesaurus (23)

A common technical resource used by some in the Natural Language Programming (NLP) community is WordNet. WordNet 3.1 Synset search provides the

---

<sup>14</sup><https://www.oed.com/search/dictionary/?scope=Entries&q=understanding>

following results (where "S : " = *ShowSynset(semantic)relations*)<sup>15</sup>

### Noun

- S: (n) understanding, apprehension, discernment, savvy (the cognitive condition of someone who understands) "he has virtually no understanding of social cause and effect"
- S: (n) agreement, understanding (the statement (oral or written) of an exchange of promises) "they had an agreement that they would not interfere in each other's business"; "there was an understanding between management and the workers"
- S: (n) sympathy, understanding (an inclination to support or be loyal to or to agree with an opinion) "his sympathies were always with the underdog"; "I knew I could count on his understanding"
- S: (n) reason, understanding, intellect (the capacity for rational thought or inference or discrimination) "we are told that man is endowed with reason and capable of distinguishing good from evil"

### Verb

- S: (v) understand (know and comprehend the nature or meaning of) "She did not understand her husband"; "I understand what she means"
- S: (v) understand, realize, realise, see (perceive (an idea or situation) mentally) "Now I see!"; "I just can't see your point"; "Does she realize how important this decision is?"; "I don't understand the idea"
- S: (v) understand, read, interpret, translate (make sense of a language) "She understands French"; "Can you read Greek?"
- S: (v) understand, infer (believe to be the case) "I understand you have no previous experience?"
- S: (v) sympathize, sympathise, empathize, empathise, understand (be understanding of) "You don't need to explain—I understand!"

### Adjective

- S: (adj) understanding (characterized by understanding based on comprehension and discernment and empathy) "an understanding friend"

As part of the search for definitions, I did ask ChatGPT for what it had to offer regarding words and concepts associated with understanding. Coupled with that, I asked for ways of ranking relatedness. Here's Chat's response:

Here's a table of words and concepts similar to, adjacent to, or related to "Understanding," organized by level of relatedness. The relatedness metric is explained after the table:

The metric used for relatedness (Rel.) is a subjective numerical measure ranging from 1 to 10, with 10 being most closely related to "Understanding" and 1 being least related. This measure is based on how directly each word or concept contributes to or is involved in the process of understanding, as well as their semantic proximity to the term "Understanding" in cognitive

<sup>15</sup><http://wordnetweb.princeton.edu/perl/webwn?c=0&sub=Change&o2=&o0=&o8=1&o1=1&o7=&o5=&o9=&o6=&o3=&o4=&i=-1&h=000000000000&s=understanding>



<b>Word</b>	<b>Explanation</b>	<b>Rel.</b>
Comprehension	Grasping the meaning, nature, or importance of; similar to understanding in terms of cognitive grasp.	9.9
Insight	The capacity to gain an accurate and deep intuitive understanding of a person or thing.	9.0
Awareness	Knowledge or perception of a situation or fact, less in-depth than understanding but related.	8.0
Knowledge	Information, understanding, or skill that you get from experience or education.	7.5
Grasp	The ability to hold or seize; in metaphorical terms, to understand concepts.	7.0
Perception	The way you think about or understand someone or something; less comprehensive than understanding.	6.5
Appreciation	Recognition and enjoyment of the good qualities of someone or something; can imply a less detailed understanding.	6.0
Familiarity	Close acquaintance with or knowledge of something; less comprehensive than full understanding.	5.5
Cognition	The mental action or process of acquiring knowledge and understanding through thought, experience, and the senses.	5.0
Interpretation	The action of explaining the meaning of something; implies an application of understanding.	4.5
Acknowledgment	Acceptance of the truth or existence of something; can be seen as a preliminary step to understanding.	4.0
Intuition	The ability to understand something immediately, without the need for conscious reasoning.	3.5
Recognition	Identification of something as having been previously seen, heard, known, etc.	3.0
Learning	The acquisition of knowledge or skills through study, experience, or being taught; related to the process of developing understanding.	2.5
Curiosity	A strong desire to know or learn something; fuels the pursuit of understanding but is not understanding itself.	2.0

Table 1: **Understanding Adjacent Terms**

and educational contexts. A higher number indicates a closer conceptual relationship or a greater degree of similarity in meaning or application.

If we were to use cosine similarity to create a table of words and concepts similar to "Understanding," the table would be based on the cosine similarity scores between the vector representation of the word "Understanding" and the vector representations of other words or concepts. This approach would require computational analysis of text data, typically using word embedding models such as Word2Vec, GloVe, or fastText.

These models represent words as high-dimensional vectors, where each dimension captures some aspect of the word's meaning based on its usage in a large text corpus. The cosine similarity between any two word vectors quantitatively reflects how similar the words are in terms of their semantic contexts in the corpus.

For practical reasons, my use of understanding will use the primary ones provided by the *OED*. These pretty much are the main senses used by most of the authoritative linguistic definition references.

**understanding**, in understand, v. To comprehend; to apprehend the meaning or import of; to grasp the idea of.

**understanding**, n. (Without article.) Power or ability to understand; intellect, intelligence.

⇔

# 1 The Big Picture

Before diving into exploring understanding in depth, to me, its worth setting the Big Picture context.

Basically, my framing approach is to explore various rabbit holes along the puzzles of existence and emergence described by the following flows:

$$\text{Nothingness} \xrightarrow{\text{Emergence of Existence}} \text{ZeroPoint Universe (ZPU)} \quad (1)$$

$$\text{ZPU} \xrightarrow{\text{Emergence of Matter}} \text{Galaxies, Star Systems, and Planets(GSP)} \quad (2)$$

$$\text{GSP} \xrightarrow{\text{Emergence of Life}} \text{Life in the Universe(Life)} \quad (3)$$

$$\text{Life} \xrightarrow{\text{Emergence of Brains}} \text{Brains} \quad (4)$$

$$\text{Brains} \xrightarrow{\text{Emergence of Minds}} \text{Minds} \quad (5)$$

$$\text{Minds} \xrightarrow{\text{Emergence of Understanding}} \text{Understanding} \quad (6)$$

$$\text{Understanding} \xrightarrow{\text{Emergence of Control}} \text{Power over the Universe} \quad (7)$$

Abstractly, one can consider emergence of the universe along *no life* and *with life* paths. We don't yet know how abundant life in the Universe is. Certainly in our little corner there *is* life since we obviously exist (notwithstanding some philosophical skepticism for the sheer clicks).

we can diagram paths as follows:

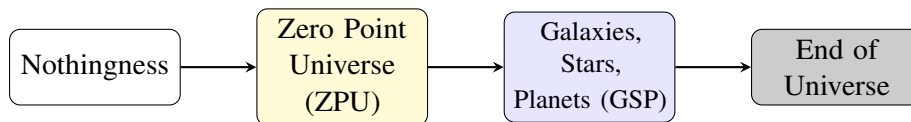


Figure 1: *No Life Path* of Universe's Progression

Figure 1 represents the highest abstraction of going from the *Nothingness* to *The End of The Universe*. The initial Nothingness and the End are not well understood. There are all kinds of speculations there. The Big Bang theory really described the parts after something started emerging, but if anyone claims they know what the prior state was before is not dealing with science.

How the Universe ends is a subject occupying physicists, cosmologists and theologians. This will not be discussed further here.

The *Zero Point Universe*(ZPU) is my label for the Universe at the instant just prior to the Big Bang. At ZPU, there's lots and lots of energy in an unstable state that "forces" the *Big Bang* to happen. The Big Bang is followed by the appearance of matter and anti-matter, and the usual progression from that to build up of elements etc. etc. culminating in the observable Universe we know with Galaxies, Stars, Planets, Interplanetary and intergalactic dust and so on. At this point, the possibilities of *forks in the road* show up. We exist in that fork that allows life to emerge and develop. We're working on figuring out on how that happens. There could well be many regions in the Universe that have no life present. There could also be lots or regions with different types of Life. Possibly there are regions that the subforks preclude development and progression exclude of any type of life. This area is full of complexity. There no agreed upon general definitions of life, although there many attempts to do so. While fascinating, a fuller discussion of these aspects would take too long and so I leave it for another time.

The primary *fork in the road* of interest to us is the one with the path we're on, the one that allowed the kind of life we know to develop. At a high level of abstraction, we can pick up the progression from the GSP stage, that culminates in Brains. On Earth we know many creatures have Brains; Wikipedia has a nice way of describing the current state of affairs "The brain is an organ that serves as the center of the nervous system in all vertebrate and most invertebrate animals"<sup>16</sup>. At a certain level of complexity, creatures with sufficiently large and complex structures start having what we call higher cognitive abilities. This yet another fascinating area. At some point brain sizes and complexity give rise to minds, which is where the focal area of our discussion.

There are probably thousands perhaps tens of thousands of books, and hundreds of thousands (if not millions) of scholarly papers that discuss *Mind*. For practical reasons, I will use the Wikipedia leading paragraph<sup>17</sup>:

The mind (adjective form: mental) is that which thinks, imagines, remembers, wills, and senses, or is the set of faculties responsible for such phenomena. The mind is also associated with experiencing perception, pleasure and pain, belief, desire, intention, and emotion. The mind can include conscious and non-conscious states as well as sensory and non-sensory experiences.

The progression from Brain to Mind and the relationships of Minds and Brains has been the topic of quite a few folks from all walks of life. From the common person to the most elevated scholars and theologians, the mind-body relationship puzzle has animated and inspired many. We are more sophisticated these days in considering complex scientific and philosophical aspects aiming to clarify how mental states, encompassing thoughts, beliefs, and consciousness, are interconnected with the physical aspects of the human body.

Historically, this puzzle has sparked extensive debate and has led to the development of various theories attempting to explain the dynamics of this relationship. *Dualism*, for example, posits that the mind and body are separate substances,

---

<sup>16</sup><https://en.wikipedia.org/wiki/Brain>

<sup>17</sup><https://en.wikipedia.org/wiki/Mind>

each existing independently yet somehow interacting. In contrast, *Physicalism* suggests that everything about the mind can be explained in terms of physical processes within the body. Meanwhile, idealism argues that reality is fundamentally mental or spiritual in nature. Psychological, medical, and neuroscientific researchers and thinkers have been exploring how subjective experiences arise from the physical processes of the brain, how mental states can affect physical health, and vice versa. We have collected and keep collecting a lot of data, especially with more accurate and higher resolution devices and technologies that measure a variety of information from single neuron level electrophysiological stimulation and recording to aggregate area measurements. We have no answers yet. Maybe some day we will. Maybe soon.

This is where we now encounter the phenomenon / process of *understanding*.

Figure 2 depicts the segment: *GSP* → *Life* → *Brains* → *Minds* of the Universe’s progression of development and emergence.

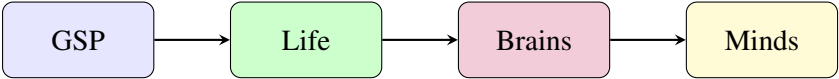


Figure 2: *Life Path* as part of Universe’s Progression

To be sure, the process of going from Nothingness to the Zero Point Universe, and then to the Big Bang, as I phrase it, especially going from nothingness to a starting point with enough energy for a is a mystery! Maybe we’ll never know the answer, then again, maybe in a million years we all have better ideas and explanations.

### 1.1 Brains to Minds to Understanding

I keep reflecting on this part of David Deutsch’s assertion:

**"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."**

To my mind, Understanding is the key for gaining power over the universe to a degree previously unimaginable. Perhaps to sufficient to satisfy as McCorduck phrased it, the *ancient wish to forge the gods*.

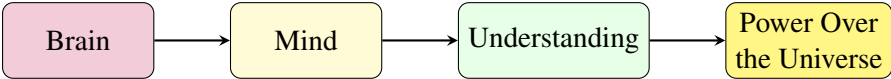


Figure 3: *Understanding* as part of the Universe’s Progression

There are others that expressed similar or adjacent opinions. Roger Penrose argued in *The Emperor’s New Mind*<sup>18</sup> and *Shadows of the Mind*<sup>19</sup> that human consciousness and understanding cannot be replicated by machines or algorithms. More generally, he held that human thought is non-algorithmic

<sup>18</sup>Penrose R. (1989). *The emperor’s new mind: concerning computers minds and the laws of physics*. Oxford University Press.

<sup>19</sup>Penrose R. (1994). *Shadows of the mind: a search for the missing science of consciousness*. Oxford University Press.

and involves elements that cannot be entirely captured by computational processes. John Searle, a philosopher known for his *Chinese Room Argument*<sup>20</sup> also argued that no computer programs could ever be said to "understand" language or anything else, in the same way humans do. He maintained that while computer programs can simulate understanding, they could not possess real understanding or intentionality.

OK, I happen to like Deutsch's formulation. Upon rereading it, several thoughts occur. He says "at present we know of nothing is capable of understanding an explanation". This formulation does not exclude the possibility of non-human entity ever being capable of understanding an explanation", (that extends also to the intentional idea of wanting one), Also, in this assertion he connects understanding and the human mind.

These consideration are mostly applied to computers and machines that perform computations. What about animals? Do animals understand things? Does a dog understand a verbal command? Do cats or elephants or dolphins or orangutans try to understand unusual events? I won't pursue understanding and explanation seeking in other animals. It is worth pondering these question though.

Regarding minds, there are varying opinions about what minds are (or if they even exist). Tentatively we can work with the commonly used notion of mind, but I offer Marvin Minsky's (MIT computer guru and prof) *Society of Mind* approach<sup>21</sup> where he posits multiple mind-agents within the brain, competing and cooperating to return a collective response of all the minds in the brain.

I also offer for consideration Erwin Schrödinger, one of Quantum Mechanics founders (the Schrödinger wave equation being one of the most important of all times). His perspectives in *Mind and Matter*<sup>22</sup> explore the concept of unity in consciousness. He proposes that the sense of individuality is an illusion and that there is a fundamental oneness connecting all conscious beings. So maybe for him, there's only a universal minds of which individuals, being physically limited organisms, can only perceive a small part they identify with their 'selves' and their consciousness.

## 1.2 From Understanding to Power Over the Universe

Figure 3 completes the depiction of Universe's progression of development and emergence. What's new here, is that the phenomenon of *understanding* is explicitly called out, and further more, identified as a step to allow entities with understanding to have *Power Over the Universe*. While this may seem as exaggeration, the question is really of scale. Right now, at our stage of human civilizational development, our impact includes mild terraforming on Earth, satellites and broadcasts sent into space and nearby planets. We've already nudged a minor asteroid, and we definitely are planning for a Near

---

<sup>20</sup>[https://en.wikipedia.org/wiki/Chinese\\_room](https://en.wikipedia.org/wiki/Chinese_room)

<sup>21</sup>Minsky, M. (1986). *The Society of Mind*. Harvard University Press. The book proposes that the mind is not a singular entity but rather a collection of smaller components, which he terms "agents." These agents, then work both independently and together to produce what we experience as thinking, feeling, and consciousness.

<sup>22</sup>Schrödinger, E. (1958). *Mind and matter*. Cambridge University Press. based on the Turner Lectures at College, Cambridge. In these lectures, Schrödinger delves into the relationship between the physical processes described by physics and the mental processes associated with consciousness and perception.

Earth Asteroid deflection technology to protect Earth (from Extinction Level Meteor strikes). Before long, we'll likely place massive sunlight capturing systems between us and the sun. We're not yet at the stage of moving our home star Sol around, but designs for that are already being talked about.

The notion of *Power Over the Universe* is a complex and unusual one. This notion deserves significant scrutiny and discussion. Hopefully I get to do that at some point. Basically, the idea here is that the Universe without Life, would develop along a path governed by basic physical laws. It progresses along the path that the Big Bang theory essentially described with the various cosmological, astrophysical we understand governed by the four fundamental forces (gravitation, electromagnetism, weak, and strong forces). Materials at high enough density and at certain temperature ranges would experience chemical phenomena, allowing various complicated molecules and crystals, and other intermediate phases to occur. However, no self replicating organic molecules and cellular structures that underlie life would develop.

In time, in such a Universe, would develop either by expanding until it cools down (Deep Freeze), or contracts (Deep Crunch), or maybe just develops to some equilibrium state where not much interesting happens.

On the Life path, namely the Universe we experience on Earth, quite a few complex developments occurred and will continue to occur as result of emergence of new types of aggregates of molecules, cells, combinations of cells, tiny and large living organisms, societies and more. All these living things modify the Universe at the molecular level as well as at the macro level in ways that an inanimate Universe could never do. There's essentially no chance that the universe, of its own, would construct a Notre Dame Cathedral<sup>23</sup>, or any cultural artifact, nor would it contain beehives or termite mounds. With humans and our creations, we've opened up the *Possibility Aperture* to allow incomprehensible large number and variety of possible 'things' to be constructed. From tangible physical things, to theories that exist in the mind (like Number Theory or Euclidean Geometry).

It is highly likely that various forms of life do possess the ability to understand some environments and phenomena, but there are definitely levels of understanding capability. The ability to design and produce mechanical birds called jet planes requires understanding that even most humans don't have.

### 1.3 But can we really have Power Over the Universe?

*Give me a place to stand and I will move the earth*  
Archimedes talking levers with Pappus of Alexandria  
Ceccarelli, Contributions of Archimedes<sup>24</sup>

Well, Archimedes of Syracuse (~ 287 – 212 BC) certainly thought he could move the Earth given the equipment. Does this count as having power of the Universe? I illustrated before that living things do have power over a very tiny part of the Universe. So the question becomes, what really counts as having *serious power*. Is it moving planets? Is it moving entire star systems? Is it

<sup>23</sup>[https://en.wikipedia.org/wiki/Notre-Dame\\_de\\_Paris](https://en.wikipedia.org/wiki/Notre-Dame_de_Paris)

<sup>24</sup>Ceccarelli, M. (2014). Contributions of Archimedes on mechanics and design of mechanisms. Mechanism and Machine Theory, 72, 86-93.

creating new stars, planets and galaxies *de novo*? How large or significant an impact does one have to have over the Universe or parts of the Universe to count as having power? Thorough discussion of this is a horse of a totally different color, and as such, I leave that to another place and time.

I do want to note that there are many who discussed potential human impact over the Universe. About 35 years ago (or maybe longer than that), I came across references to the work of Teilhard de Chardin's *The Phenomenon of Man*<sup>25</sup>. The work was originally published in 1940 brought up many fascinating notions ... especially *noosphere* and the idea of cosmic destiny culminating in what he terms the Omega Point. De Chardin, a Jesuit priest and an able scientist (paleontologist) articulated a powerful visions of a human driven continually improving arch of history for the entire universe. De Chardin's vision is christological vision of the future. I don't share his theological perspective, I am fascinated though by the argument of man's capacity and ability to influence the Universe.

A more hard-nosed physicalist perspective was provided by Valentin Turchin, an eminent computer scientist and cybernetician<sup>26</sup>. Turchin builds on de Chardin's perspectives and describes a sequence of sociotechnical processes and emergent evolution of complexities and richness from the point of cybernetics and *MetaSystems Transitions (MSTs)* described and elaborated in his seminal book *The phenomenon of science*<sup>27</sup>. Turchin's treatment shifts from the spiritual to nuts and bolts technical considerations and provides a grand cybernetic meta-theory of universal evolution.

For Turchin, MetaSystem Transitions are a result of evolutionary processes in which higher levels of control emerge in system structure and function. His focus is on hierarchies of control, and emergence of new hierarchical levels. Turchin makes this assertion

THE UNIVERSE IS EVOLVING. The organization of matter is constantly growing more complex. This growing complexity occurs through metasytem transitions from which new levels of organization emerge which are levels of the control hierarchy. The inorganic world, plants, animals, the human being—such has been the course of evolution on our planet, and as far as we know this is the greatest advance which has been made in the part of space that surrounds us. It also seems highly probable that the human being is the crown of evolution of the entire cosmos.

He proceeds to make the following claims:

The appearance of the human being marks the beginning of the Age of Intellect, when the leading force of development becomes conscious human creativity and the highest level of organization is the culture of human society. In its development culture generates the next level of the hierarchy within itself. This is critical thinking which, in its turn, gives rise to modern science, constructing models of reality using sign systems. These are new models; they did not and could not exist in the minds of individual human beings outside of civilization and culture, and they enlarge human power over nature colossally. They make up the

---

<sup>25</sup>Teilhard de Chardin P. (1959). *The phenomenon of man* (Rev. English). Harper & Row.

<sup>26</sup>[https://en.wikipedia.org/wiki/Valentin\\_Turchin](https://en.wikipedia.org/wiki/Valentin_Turchin)

<sup>27</sup>Turchin, V. F. (1977). *The phenomenon of science*, Columbia University Press.



continuously improving and developing super-brain of the super-being which is humanity as a whole. Thus, science is the highest level of the hierarchy in the organization of cosmic matter. It is the highest growth point of a growing tree, the leading shoot in the evolution of the universe. This is the significance of the cosmic phenomenon of science as a part of the phenomenon of man.

Specifically, he calls out:

- Power over nature
- Science . . . the highest level of the hierarchy in the organization of cosmic matter

Turchin devotes considerable time and space to discuss ethics, morality, values, society, language, and various philosophical considerations. What stands out to me is his assertion and confidence in the emergence of human control over nature, and perhaps the entire Universe. In his words:

The assertion of the continuous development and evolution of the universe is the most important general truth established by science. Everywhere we turn we observe irreversible changes subordinate to a majestic general plan or to the basic law of evolution, which manifests itself in the growing complexity of the organization of matter. Reason emerges on Earth as a part of this plan. And although we know that the sphere of human influence is a tiny speck in the cosmos still we consider the human being the crown of nature's creation. Experience in investigating the most diverse developing systems shows that a new characteristic appears first in a small space but, thanks to the potential enclosed in it, engulfs a maximum of living, space over time and creates the springboard for a new, higher level of organization. Therefore we believe that a great future awaits the human race, surpassing everything that the boldest imagination can conceive.

In 1995 I started reflecting seriously on the idea of MetaSystem Transitions and their implications. I wrote a paper on Metasystem Transitions, memes, and *Cybernetic Immortality*<sup>28</sup> Now almost thirty years later, especially with the accelerating Advances in AI, my approach is to consider the MST, evolution, and emergence discussions more carefully. What I find now, is that the *human ability to understand* is perhaps the most pivotal. It is the understanding, and the ability to employ understanding in tangible ways, that really make the most significant contributions to future progress and control over nature and the Universe.

To conclude this section, let me repeat the view that there are certain significant steps in emergence from *Nothingness* to *Power over the Universe*. *Understanding* is a distinct process that involves *Brains* and *Minds* and as I describe it, it is an emergent phenomenon from those. Also to be sure, I am not ignoring the fact that there are collective aspects to increasing knowledge and understanding that involve many humans, perhaps all of society and civilization. These considerations are indeed appropriate, but for the rest of the paper, I am focusing on considering the abstraction of Understanding as it involves current developments in Large Language Models, in particular those associated with GPT-4.

---

<sup>28</sup>Moritz, E. (1995). Metasystem transitions, memes, and cybernetic immortality. *World Futures: Journal of General Evolution*, 45(1-4), 155-171. <https://www.tandfonline.com/doi/abs/10.1080/02604027.1995.9972558>

## 2 Chatting with Chat

This section records my 'conversations' with ChatGPT<sup>29</sup>. I used the paid version known as GPT-4 (sometimes referred to as GPT4). I refer to it as *Chat*. Chat states its training includes information up until December 2023.

The conversations here, for the most part, take the form below. I pose a question or query. Formally this is known as a *prompt*. I then record the verbatim responses provided by Chat; occasionally I may skip including the summaries if they are too repetitious or not very useful. To set the baseline, I am asking Chat about itself. If you're not yet familiar with Chat, *User* refers to the one entering the prompts or questions. I use *italics* to highlight my prompts. When it's obvious who the user is (i.e. by use of italics, I sometimes drop the 'User' label)

Mechanics of Transcription: Chat's output is provided in Markdown format. I created a translator that converts Markdown to  $\LaTeX$  using Pandoc, running the  $\LaTeX$  compiler results produces the pdf you're seeing.

With that, we now jump into the discussion with Chat. My editorializing will use appear either as side notes or in some box. Sometimes to clarify that a segment of discussion has ended, and a new set of questions is starting, I use the double harpoon symbol ( $\Leftrightarrow$ ).

$\Leftrightarrow$

*User: describe yourself*

I am ChatGPT, a large language model developed by OpenAI, based on the GPT-4 architecture. My design allows me to process and generate human-like text based on the input I receive. I can assist with a wide range of tasks, including answering questions, providing explanations, offering advice, generating creative content, and more. I operate based on patterns and information from a vast dataset, which includes books, articles, and websites, up to my last update.

My capabilities are rooted in machine learning algorithms, and while I strive to provide accurate and helpful responses, I don't possess personal experiences or emotions. I'm here to help answer questions, facilitate learning, and engage in informative conversations based on the knowledge I've been trained on. I'm also designed to respect privacy and confidentiality and to provide information and assistance within the guidelines set by OpenAI.

$\Leftrightarrow$

*Please provide some more granular details about your GPT-4 architecture and the concept of GPT-4*

The GPT-4 (Generative Pre-trained Transformer 4) architecture is an advanced iteration in the series of transformer-based models developed by OpenAI. While specific technical details and improvements can vary from one iteration to the next, the GPT-4 model builds upon the foundational concepts established by its predecessors, incorporating advancements in machine learning, natural language processing, and deep learning. Here's a more granular look at the GPT-4 architecture and concept:

---

<sup>29</sup>located at <https://chat.openai.com/>

### 2.0.1 Core Concepts and Structure:

1. **Transformer Architecture:** Like its predecessors, GPT-4 is based on the transformer architecture, which is known for its self-attention mechanism. This allows the model to weigh the importance of different words in the input data relative to each other and to capture the context of words in a sentence or document efficiently.
2. **Large-scale Pre-training:** GPT-4 undergoes extensive pre-training on a diverse and extensive dataset comprising text from the internet, books, articles, and other sources. This pre-training involves learning to predict the next word in a sequence given the previous words, enabling the model to develop a deep understanding of language patterns, grammar, facts, and concepts.
3. **Fine-tuning and Task Agnosticism:** After pre-training, the model can be fine-tuned for specific tasks, although GPT-4 is designed to be task-agnostic at its core. This means it can perform a wide range of tasks without needing task-specific training, from language translation and question-answering to content creation and code generation.

### 2.0.2 Advancements in GPT-4:

1. **Scale:** GPT-4 is expected to be larger than its predecessor, GPT-3, in terms of the number of parameters, the size of the dataset it is trained on, and its computational power. This increase in scale typically leads to improved understanding and generation capabilities, although it also raises concerns regarding energy consumption and computational efficiency.
2. **Improved Contextual Understanding:** GPT-4 aims to provide better contextual understanding and coherence over longer stretches of text, addressing one of the limitations of earlier models.
3. **Reduced Biases and Errors:** Efforts have been made to reduce biases and errors in GPT-4's responses compared to earlier versions. This involves more sophisticated training techniques, better dataset curation, and mechanisms to handle sensitive or controversial topics more delicately.
4. **Multimodal Capabilities:** While primarily known for text generation, GPT-4 and its associated technologies might incorporate or be adapted to understand and generate other types of data, such as images, audio, and possibly even video, enhancing its multimodal capabilities.
5. **Interactivity and Customization:** GPT-4 offers improved interactivity, allowing for more dynamic and context-aware conversations. It also provides options for customization, enabling users to tailor the model's responses to specific needs or preferences.

### 2.0.3 Limitations and Ethical Considerations:

Despite its advancements, GPT-4, like all AI models, has limitations. It can still generate incorrect or misleading information, exhibit biases present in its training data, and lack true understanding or consciousness. Ethical considerations, such as privacy, misinformation, and the impact on jobs, remain critical areas of discussion and research.

In summary, GPT-4 represents a significant step forward in natural language processing and artificial intelligence. Its architecture and training enable a wide range of applications, from creative writing to technical problem-solving, while also presenting challenges and questions that the field continues to address.

⇔

## 2.1 Starting Line: Deutsch's remarks on Understanding

*discuss "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." in granular detail*

This statement captures the essence of what distinguishes human cognition from artificial intelligence and the instinctual behavior observed in animals. It underscores the complexity and uniqueness of human understanding, a trait that transcends mere data assimilation or reactive behaviors. Let's break it down into its components for a detailed analysis:

### 2.1.1 Understanding as a Higher Function

- **Human Mind and Brain:** The phrase highlights the human mind and brain's role in understanding, suggesting that this capability is not just a product of the physical brain but also involves the abstract functions of the mind. This implies a combination of neural activity and cognitive processes that go beyond the biological hardware.
- **Unique Function:** The uniqueness of understanding among cognitive functions suggests it's not only about processing information but involves deeper layers of comprehension, interpretation, and the application of knowledge in a way that is currently beyond other entities.

### 2.1.2 Comparison with Other Systems

- **Animals' Brains:** While animals can learn, remember, and react to their environment, their abilities are often driven by instinct, conditioning, or simple cognitive processes. The statement implies that animals do not seek or require explanations in the same way humans do, highlighting a fundamental difference in the nature of cognitive processing.
- **Computers and Machines:** These systems can process vast amounts of data, perform complex calculations, and execute tasks based on programmed instructions. However, they lack the ability to understand context, meaning, or the nuances of human thought and language beyond their coded algorithms. The distinction here is between executing pre-defined tasks and genuinely understanding the why behind actions or facts.

### 2.1.3 The Nature of Understanding

- **Capability of Understanding an Explanation:** This points to a cognitive ability to not only receive information but to comprehend its

underlying principles, context, and implications. Understanding an explanation involves grasping the cause-and-effect relationships, abstract concepts, and the intention behind the communicated information.

- **Desire for Explanation:** Beyond the capability to understand, humans are unique in their curiosity and desire to seek explanations. This desire drives scientific inquiry, philosophical thought, and personal reflection. It's a fundamental aspect of learning, growth, and the pursuit of knowledge.

#### 2.1.4 Implications and Questions

This statement opens up several lines of inquiry and reflection:

- **Nature of Consciousness:** What is it about human consciousness that enables understanding, and how does it emerge from the neural processes within the brain?
- **Artificial Intelligence:** Can AI ever achieve a level of understanding comparable to humans, or will it remain a tool that simulates aspects of human cognition without true comprehension?
- **Evolution of Understanding:** How did the ability and desire for understanding evolve in humans, and what does this tell us about the nature of intelligence and consciousness?

The uniqueness of human understanding, as outlined in the statement, raises profound questions about the intersection of biology, technology, and philosophy. It challenges us to consider what it means to truly understand something and what this ability means for our place in the natural world.

⇔

*User: provide a thorough discussion of scholars' writing about human understanding*

Human understanding has been a central topic of inquiry across various disciplines, including philosophy, psychology, cognitive science, and artificial intelligence research. Scholars from these fields have approached the concept from different angles, focusing on its mechanisms, implications, and the distinctions between human cognition and other forms of intelligence. Here's an overview of key scholarly contributions to the discussion of human understanding:

#### 2.1.5 Philosophical Perspectives

- **Immanuel Kant:** In his Critique of Pure Reason, Kant delves into how human understanding is possible, introducing the concept of a priori knowledge and the categories of understanding. He argues that human understanding shapes our experience of the world through innate structures of the mind, emphasizing the active role of the mind in constructing reality.
- **John Locke:** Locke's An Essay Concerning Human Understanding explores the origin of human thoughts and the nature of knowledge. He posits that all knowledge originates from experience, through the senses,

challenging the notion of innate ideas and emphasizing the importance of empirical evidence in understanding.

- **Gottfried Wilhelm Leibniz:** In contrast to Locke, Leibniz suggests in his works, such as *New Essays on Human Understanding*, that the mind possesses innate principles or ideas that play a crucial role in understanding the world, highlighting the complexity of human cognition.

### 2.1.6 Psychological and Cognitive Science Views

- **Jean Piaget:** A developmental psychologist, Piaget's theory of cognitive development outlines how children develop understanding through stages, emphasizing the role of active learning and the construction of knowledge through interaction with the environment.
- **Daniel Kahneman:** Kahneman's work, particularly in *Thinking, Fast and Slow*, explores the dual-process theory of the mind. He distinguishes between two types of thinking: fast, intuitive thought, and slow, rational thought, shedding light on how humans understand and make decisions.
- **Steven Pinker:** In *How the Mind Works* and other writings, Pinker investigates the nature of human thought and understanding from an evolutionary psychology perspective. He explores how the brain's structure and function have been shaped by evolutionary pressures to solve problems related to survival and reproduction.

### 2.1.7 Artificial Intelligence and Computational Models

- **Alan Turing:** Turing's pioneering work laid the foundation for computational thinking and artificial intelligence. His concept of the Turing Test challenges the distinction between human and machine understanding, proposing a criterion for a machine's ability to exhibit intelligent behavior equivalent to, or indistinguishable from, that of a human.
- **Marvin Minsky:** As a co-founder of the MIT Artificial Intelligence Laboratory, Minsky's work on frames and knowledge representation in *The Society of Mind* offers insights into how machines could mimic human understanding by structuring knowledge in interconnected modules.
- **John Searle:** In his famous Chinese Room Argument, Searle argues against the possibility of computers truly understanding in the same way humans do. He suggests that computational processes can simulate understanding but lack the subjective experience, or qualia, associated with human cognition.

### 2.1.8 Interdisciplinary Approaches

- **Noam Chomsky:** Chomsky's theory of Universal Grammar proposes that the ability to acquire language is innate to humans, reflecting a unique aspect of human understanding related to the structure and use of language. This has implications for understanding the cognitive processes underlying language acquisition and use.
- **Antonio Damasio:** Damasio's research in neuroscience, particularly in "Descartes' Error," emphasizes the role of emotion and the body

in cognition and decision-making, challenging traditional dichotomies between emotion and rationality in understanding.

These scholars, among many others, contribute to a multifaceted view of human understanding, highlighting its complexity and distinguishing features. The ongoing dialogue between disciplines continues to enrich our comprehension of how humans understand the world, the unique capabilities of the human mind, and the potential for creating machines that might one day emulate these capabilities.

⇔

## 2.2 On Artificial Intelligence and Understanding

*User: Discuss, in granular detail, the State of the Art of research and development of AI technologies, including GPTs and LLMs, that focus on enabling and increasing AI 'understanding' abilities.*

The state of the art in AI research and development, especially in areas focusing on Generative Pre-trained Transformers (GPTs) and Large Language Models (LLMs), has seen remarkable progress in recent years. These advancements aim to enhance AI systems' ability to understand, process, and generate human-like text, moving closer to a form of understanding that mirrors human cognitive abilities. Let's delve into the granular details of the current landscape, including methodologies, key breakthroughs, challenges, and future directions.

### 2.2.1 GPTs and LLMs: Foundations

- **Generative Pre-trained Transformers (GPTs):** These are a class of language models that use the transformer architecture for generating text. The model is first pre-trained on a large corpus of text data in an unsupervised manner to learn the statistical properties of language. It's then fine-tuned for specific tasks. GPT-3, developed by OpenAI, is one of the most prominent examples, known for its ability to generate coherent and contextually relevant text across a wide range of domains.
- **Large Language Models (LLMs):** LLMs refer to any substantial neural network models trained on vast datasets to understand and generate language. This category includes models like BERT (Bidirectional Encoder Representations from Transformers) and its successors, which are designed to comprehend the context of words in a sentence more effectively than their predecessors.

### 2.2.2 Increasing AI 'Understanding' Abilities

1. **Contextual Understanding:** Modern AI models have made significant strides in grasping the context of text. This is achieved through attention mechanisms and transformer architectures that analyze the relationships between words in a sentence or across sentences, enabling a more nuanced interpretation of language.
2. **Knowledge Integration:** Efforts are being made to integrate external knowledge bases with LLMs to enhance their understanding. This involves techniques like entity linking, where entities in text are linked

to a knowledge base, providing the model with access to a broader context beyond the input text.

3. **Multimodal Models:** Research into multimodal models, which process and understand information from various data types (e.g., text, images, audio), represents a leap towards more holistic forms of AI understanding. These models, like OpenAI's CLIP and DALL·E, demonstrate an ability to grasp complex concepts that span different forms of data, showcasing an advanced level of comprehension.
4. **Commonsense Reasoning:** Enhancing AI's commonsense reasoning abilities remains a critical area of research. Models are being trained to understand and apply commonsense knowledge to their outputs, allowing them to make inferences and decisions that align more closely with human reasoning.
5. **Continual Learning:** Traditional models require extensive retraining to update their knowledge. Continual learning approaches aim to enable models to learn from new data continuously without forgetting previously learned information, mimicking the human ability to accumulate knowledge over time.

### 2.2.3 Challenges and Future Directions

- **Understanding vs. Simulation:** Despite their advancements, current AI models primarily simulate understanding through statistical pattern recognition. True comprehension, involving self-awareness, consciousness, and the ability to experience, remains beyond their reach.
- **Bias and Fairness:** AI models can perpetuate or even amplify biases present in their training data. Addressing these biases to ensure fair and ethical outcomes is an ongoing challenge.
- **Energy Efficiency:** The computational resources required for training and running state-of-the-art LLMs are immense. Research into more energy-efficient models and algorithms is critical for sustainable development.
- **Explainability:** As models become more complex, ensuring their decisions and processes are transparent and interpretable to humans is increasingly challenging but necessary for trust and ethical considerations.

### 2.2.4 Conclusion

The state of the art in AI, particularly in the realm of GPTs and LLMs, is pushing the boundaries of what machines can understand. Through innovations in architecture, training methodologies, and the integration of multimodal data and external knowledge, AI systems are achieving unprecedented levels of linguistic and conceptual processing. However, the journey towards true AI understanding—encompassing self-awareness, ethical reasoning, and genuine comprehension—remains a long-term challenge, driving forward the frontiers of research and development in the field.



## 2.3 AI SOTA Limitations

*User: discuss the limitations of AI and SOTA models to actually understand*

While state-of-the-art (SOTA) models in artificial intelligence (AI), such as Generative Pre-trained Transformers (GPTs) and other Large Language Models (LLMs), demonstrate remarkable capabilities in processing and generating human-like text, their ability to understand in the human sense is subject to significant limitations. These limitations stem from the fundamental nature of how AI models are designed, trained, and operate. Here's a detailed discussion of the key limitations:

### 2.3.1 Lack of True Comprehension

- **Simulating Understanding:** AI models, including the most advanced ones, simulate understanding by predicting the likelihood of sequences of words based on patterns learned from vast datasets. However, this process lacks the conscious awareness and genuine comprehension that humans experience. AI models do not possess intentions, beliefs, desires, or consciousness, which are critical elements of true understanding.

### 2.3.2 Contextual and Nuanced Interpretation

- **Shallow Contextual Grasp:** Despite advances in contextual understanding through transformer architectures, AI's grasp of context remains fundamentally shallow. Models can match patterns and generate plausible responses, but they do not deeply understand the cultural, historical, or deeply personal nuances that humans naturally integrate into their comprehension of language and concepts.

### 2.3.3 Commonsense Reasoning and Knowledge

- **Limited Commonsense Reasoning:** AI struggles with commonsense reasoning, which involves making inferences about the world that are obvious to humans but not explicitly stated in the data. Efforts to incorporate commonsense knowledge into AI systems are ongoing, but the ability to seamlessly apply this knowledge in a wide range of scenarios, as humans do, remains limited.

### 2.3.4 Learning and Adaptation

- **Static Knowledge Base:** Once trained, the knowledge base of an AI model remains largely static unless it is specifically updated or retrained with new data. This contrasts with human learning and adaptation, where new information is continuously integrated into our understanding, and misconceptions can be corrected through new experiences or information.

### 2.3.5 Ethical and Moral Understanding

- **Absence of Ethical Reasoning:** AI models do not possess ethical reasoning or moral understanding. They can simulate advice or actions based on patterns in data but cannot genuinely grasp ethical principles or the

moral implications of actions. This limitation is particularly concerning when AI is applied in contexts requiring nuanced ethical judgments.

### 2.3.6 Emotional Intelligence

- **Lack of Emotional Understanding:** Emotional intelligence, a key component of human understanding, involves recognizing, interpreting, and responding to emotions in oneself and others. AI can simulate empathetic responses based on training data but does not truly understand or experience emotions, limiting its ability to fully comprehend human-centric texts and interactions.

### 2.3.7 Creativity and Innovation

- **Constrained Creativity:** While AI models can generate novel content by recombining elements in new ways, their creativity is fundamentally constrained by the patterns and data they have been trained on. True creativity involves breaking away from existing patterns to create something genuinely new and original, a feat AI is currently unable to achieve autonomously.

### 2.3.8 Future Directions

Addressing these limitations requires advancements not only in AI technology but also in our understanding of consciousness, cognition, and the essence of human understanding. Future research may explore more sophisticated models of knowledge representation, novel training paradigms that enable continuous learning and adaptation, and interdisciplinary approaches that integrate insights from cognitive science, philosophy, and ethics.

Despite their limitations, SOTA AI models continue to push the boundaries of what machines can do, offering tools and insights that complement human abilities. However, the quest to bridge the gap between simulating understanding and achieving a form of AI that possesses genuine comprehension, emotional intelligence, and ethical reasoning remains a profound challenge for the future of artificial intelligence research.

⇔

## 2.4 Evidence of Understanding by SOTA AI

*User: Discuss in granular detail what would constitute evidence of understanding by SOTA AI*

Evidence of understanding by state-of-the-art (SOTA) artificial intelligence (AI), particularly in the context of models like Generative Pre-trained Transformers (GPTs) and Large Language Models (LLMs), would require a multifaceted approach to evaluation. It involves assessing the AI's performance across several dimensions that collectively could indicate a form of understanding beyond mere pattern recognition. Here are detailed criteria that would constitute evidence of understanding by SOTA AI:

### 2.4.1 Contextual Reasoning and Adaptability

- **Granular Contextual Understanding:** The AI demonstrates the ability to grasp the context at a nuanced level, including understanding the implications of cultural, historical, and situational contexts, and adjusting its responses accordingly.
- **Dynamic Context Adaptation:** The ability to adapt its understanding and responses based on changing contexts or new information introduced in the course of an interaction, resembling how humans can shift their understanding as conversations evolve.

### 2.4.2 Commonsense Reasoning

- **Applying Commonsense Knowledge:** The AI consistently applies commonsense knowledge to interpret, infer, and generate responses that align with what a reasonable person would consider obvious or typical in a given situation.
- **Commonsense Problem-Solving:** Demonstrating the capability to solve problems or answer questions by applying commonsense reasoning, showing an understanding of the physical world, social norms, and basic logic.

### 2.4.3 Deep Semantic Understanding

- **Abstract Concept Interpretation:** The ability to understand and manipulate abstract concepts, metaphors, and idiomatic expressions, translating them into coherent, contextually appropriate responses.
- **Disambiguation:** Effectively handling linguistic ambiguity, such as homonyms or syntactically complex sentences, based on the broader context, indicating a deeper level of linguistic and semantic understanding.

### 2.4.4 Emotional Intelligence

- **Empathetic Responses:** Exhibiting an understanding of human emotions through appropriate and context-sensitive responses, suggesting an awareness of the emotional content of human interactions.
- **Emotionally Congruent Generation:** The ability to generate content that not only acknowledges the emotional tone of the input but also appropriately matches or responds to it, simulating a form of emotional intelligence.

### 2.4.5 Ethical and Moral Reasoning

- **Ethical Decision Making:** Showing the capacity for ethical reasoning by weighing potential actions against a set of moral principles or societal norms and choosing actions that adhere to these ethical guidelines.
- **Understanding Moral Implications:** Demonstrating an understanding of the moral implications of actions or statements, including the ability to predict potential ethical concerns or conflicts arising from certain scenarios.

#### 2.4.6 Creativity and Innovation

- **Novel Problem-Solving:** Generating solutions to problems that are novel and not directly inferable from the training data, indicating an ability to think outside the box.
- **Original Content Creation:** The creation of original, creative content (e.g., stories, art, theories) that goes beyond recombining existing information and shows signs of a novel thought process.

#### 2.4.7 Continuous Learning and Adaptation

- **Incorporating New Information:** The capacity to learn from new information or feedback in real-time, adjusting future responses and understanding based on this continuous learning process.
- **Self-Correction:** Identifying and correcting its own errors without external intervention, demonstrating an ability to reflect on and learn from mistakes.

#### 2.4.8 Evaluation and Measurement

- **Benchmarks and Metrics:** Developing and applying sophisticated benchmarks and metrics designed to measure these aspects of understanding in a quantitative and qualitative manner.
- **Interdisciplinary Validation:** Engaging experts from various fields, including psychology, linguistics, ethics, and cognitive science, to validate the AI's performance and its alignment with human-like understanding.

Evidence of understanding in SOTA AI would thus require not just technical proficiency in language processing but an integration of cognitive, emotional, ethical, and creative capabilities that mimic the breadth and depth of human intelligence. Achieving and measuring such understanding poses significant challenges but represents a crucial frontier in AI research and development.

⇔

### 2.5 Objective Evaluation Benchmarks

*Elaborate on what would constitute highly accepted objective benchmarks*

Highly accepted objective benchmarks for evaluating the understanding abilities of state-of-the-art (SOTA) artificial intelligence (AI), including Generative Pre-trained Transformers (GPTs) and Large Language Models (LLMs), would need to be comprehensive, rigorous, and reflective of a wide range of cognitive, emotional, and ethical capabilities. These benchmarks would serve as standardized tests to assess AI systems' performance across various dimensions of understanding. Here's an elaboration on what such benchmarks might include:

#### 2.5.1 Linguistic and Semantic Understanding

- **GLUE (General Language Understanding Evaluation) and Super-GLUE:** These benchmarks assess a model's ability to understand and generate language through tasks like natural language inference, question answering, and sentiment analysis. They are designed to measure

how well an AI system can process and interpret the nuances of human language.

- **Winograd Schema Challenge:** A test of commonsense reasoning and natural language understanding that requires the model to resolve ambiguities in sentences, which is a critical aspect of human-like understanding.

### 2.5.2 Commonsense and Reasoning

- **CommonsenseQA:** A question-answering benchmark for evaluating the AI's ability to answer questions that require commonsense knowledge outside of the specific information provided in the text.
- **PIQA (Physical Interaction QA):** Focuses on the AI's understanding of physical world interactions, assessing its ability to predict the outcomes of physical processes or the use of everyday objects.

### 2.5.3 Ethical and Moral Reasoning

- **ETHICS dataset:** A benchmark designed to evaluate models on ethical reasoning, including justice, deontology, virtue ethics, utilitarianism, and commonsense moral intuitions. This benchmark assesses whether AI can make morally informed decisions or understand the ethical implications of various actions and statements.

### 2.5.4 Emotional Intelligence

- **EmoContext:** Aims to evaluate the AI's ability to understand emotional context in text conversations, assessing its capacity to recognize and appropriately respond to emotional cues.
- **GoEmotions:** A dataset for fine-grained emotion classification, which could be used to benchmark the AI's ability to discern and generate a wide range of human emotions accurately.

### 2.5.5 Creativity and Novel Problem-Solving

- **Diverse and Novel Task Performance:** Establishing benchmarks that require AI to solve problems or create content in areas not directly covered in its training data. This could involve generating novel artworks, stories, or inventive solutions to theoretical problems.
- **Adversarial Creativity Tasks:** Challenges where AI must respond to dynamically generated tasks that require creative problem-solving, ensuring the AI cannot rely on memorized solutions.

### 2.5.6 Continuous Learning and Adaptation

- **Lifelong Learning Benchmarks:** Tests that measure the AI's ability to learn continuously from new data without forgetting previously learned information, simulating the human capacity for ongoing learning and adaptation.

- **Dynamic Adaptation Tests:** Scenarios that require the AI to adjust its responses or strategies based on changing information or feedback, evaluating its flexibility and adaptability.

### 2.5.7 Implementation and Validation

For these benchmarks to be highly accepted and objective, several criteria must be met:

- **Transparency and Accessibility:** The benchmarks should be openly accessible to the research community, with a clear description of the metrics, datasets, and evaluation protocols.
- **Diverse and Inclusive Data:** The datasets used in these benchmarks should be diverse and inclusive, representing a wide range of languages, cultures, and contexts to ensure the AI systems are evaluated on their ability to understand and interact with a broad spectrum of human experiences.
- **Interdisciplinary Expertise:** Development and validation of these benchmarks should involve interdisciplinary expertise, including linguists, ethicists, psychologists, and domain-specific experts, to ensure the benchmarks accurately reflect human-like understanding.

Highly accepted objective benchmarks in AI understanding are crucial for driving progress, enabling fair comparisons across different AI systems, and guiding research towards models that truly mimic human cognitive, emotional, and ethical capabilities.

⇒

*User: Describe the Winograd Schema Challenge in granular detail. Also discuss progress and SOTA of this challenge*

The Winograd Schema Challenge (WSC) is a test of artificial intelligence that aims to evaluate a system's understanding of human language beyond simple pattern recognition, focusing on its ability to resolve ambiguities in natural language text. It was proposed by Hector Levesque, a computer scientist, as an alternative to the Turing Test, offering a more specific and clear-cut challenge to assess AI's understanding capabilities. The WSC is designed to probe the AI's commonsense reasoning and ability to interpret context, which are critical for genuine language comprehension.

### 2.5.8 Structure of the Winograd Schema Challenge

- **Winograd Schemas:** A Winograd Schema consists of a pair of sentences that differ in only one or two words and contain an ambiguous pronoun. The challenge is to determine the referent of the pronoun, which requires understanding the sentence's context and applying commonsense knowledge. The correct interpretation hinges on a single word or phrase that changes between the two sentences in the pair.
- **Example:** Consider the following pair of sentences where the ambiguous pronoun it needs to be resolved:
  - The city councilmen refused the demonstrators a permit because they feared violence.

- The city councilmen refused the demonstrators a permit because they advocated violence.

In the first sentence, they refers to the city councilmen, while in the second sentence, they refers to the demonstrators. Determining the correct referent requires understanding the causal relationship implied by the sentences.

### 2.5.9 Criteria for Challenge Problems

- **Binary Choice:** Each schema presents a question with a binary choice, forcing the AI to choose between two potential referents for the ambiguous pronoun.
- **Commonsense Knowledge and Reasoning:** Successfully answering a question requires the application of commonsense knowledge about the world and understanding the subtleties of human language.
- **Resistance to Pre-programmed Solutions:** The schemas are designed to be resistant to solutions based on keyword matching or statistical correlations in large text corpora, aiming to ensure that successful performance indicates a deeper level of understanding.

### 2.5.10 Progress and State of the Art

- **Early Performance:** Initially, AI systems struggled with the Winograd Schema Challenge, demonstrating that resolving the type of ambiguity presented in these schemas requires a level of understanding and reasoning that goes beyond the capabilities of most early models.
- **Advancements in AI:** Over time, there have been significant advancements in natural language processing (NLP) technologies, particularly with the development of Large Language Models (LLMs) like GPT-3. These models have shown improved performance on tasks similar to the WSC, thanks to their ability to process and generate human-like text based on vast amounts of training data.
- **Limitations:** Despite these advancements, achieving consistent success on the WSC remains challenging. AI systems can sometimes perform well on individual schemas through pattern recognition or leveraging statistical biases in the training data, but consistent and reliable understanding across a broad range of schemas is difficult. This suggests that while AI has made progress in language processing, genuine understanding and commonsense reasoning are still areas requiring further research.
- **Current State:** The current state of the art in AI, with respect to challenges like the WSC, involves a combination of advanced NLP models, knowledge integration (linking AI systems to structured knowledge bases), and specialized reasoning algorithms. However, no AI system has yet demonstrated the level of understanding and flexible commonsense reasoning that the WSC aims to test, indicating that this remains an open and active area of research.

The Winograd Schema Challenge continues to serve as a critical benchmark for evaluating AI's progress towards human-like understanding. It highlights

the importance of commonsense reasoning and deep semantic understanding in AI development, pushing researchers to explore new approaches that move beyond statistical learning towards models that can genuinely interpret and engage with the complexity of human language.

⇐

### 2.5.11 GLUE and SuperGLUE: An Overview

*Describe GLUE and SuperGLUE in granular detail. Also discuss progress and SOTA of this challenge*

The General Language Understanding Evaluation (GLUE) and its more challenging successor, SuperGLUE, are benchmarks designed to evaluate and promote the development of advanced natural language understanding (NLU) systems. These benchmarks consist of a collection of tasks that test a wide range of linguistic abilities, including sentence-level and word-level understanding, inference, and reasoning. The aim is to push the boundaries of what AI can achieve in comprehending human language.

### 2.5.12 GLUE Benchmark

Introduced in 2018, GLUE is a collection of nine diverse tasks that cover various aspects of language understanding:

1. **CoLA (The Corpus of Linguistic Acceptability)**: Tests whether a sentence is grammatically correct.
2. **SST-2 (The Stanford Sentiment Treebank)**: Requires the model to determine if a sentence's sentiment is positive or negative.
3. **MRPC (Microsoft Research Paraphrase Corpus)**: Asks the model to judge whether two sentences are semantically equivalent (paraphrases of each other).
4. **STS-B (Semantic Textual Similarity Benchmark)**: Involves scoring the similarity of two sentences on a scale from 1 to 5.
5. **QQP (Quora Question Pairs)**: Requires determining whether two questions asked on Quora are semantically equivalent.
6. **MNLI (Multi-Genre Natural Language Inference)**: Involves reading a premise and then determining whether a hypothesis is true (entailment), false (contradiction), or undetermined (neutral).
7. **QNLI (Question Natural Language Inference)**: Converts the Stanford Question Answering Dataset into a problem of determining whether a context sentence contains the answer to a question.
8. **RTE (Recognizing Textual Entailment)**: Similar to MNLI but with a smaller dataset and simpler texts.
9. **WNLI (Winograd Schema Challenge)**: A simplified version of the Winograd Schema Challenge, focused on pronoun disambiguation.



### 2.5.13 SuperGLUE Benchmark

Launched as a follow-up to GLUE, SuperGLUE was designed to be a more difficult set of challenges, reflecting the rapid advancements in NLU technology. It includes eight tasks:

1. **BoolQ (Boolean Questions)**: Requires answering a yes/no question given a short passage.
2. **CB (CommitmentBank)**: A three-way classification task to evaluate understanding of textual entailment.
3. **COPA (Choice of Plausible Alternatives)**: A causal reasoning task where the model chooses between two possible causes or effects.
4. **MultiRC (Multi-Sentence Reading Comprehension)**: Asks questions about a paragraph where answers may require drawing information from multiple sentences.
5. **ReCoRD (Reading Comprehension with Commonsense Reasoning Dataset)**: A reading comprehension task that requires commonsense reasoning, with multiple-choice answers.
6. **RTE (Recognizing Textual Entailment)**: This is carried over from GLUE but remains a challenging entailment task.
7. **WiC (Words in Context)**: Tests word sense disambiguation in context.
8. **WSC (Winograd Schema Challenge)**: A more difficult version compared to the one in GLUE, focusing on coreference resolution in complex scenarios.

### 2.5.14 Progress and State of the Art

- **Advancements**: Since the introduction of GLUE and SuperGLUE, there have been significant advancements in NLU capabilities, driven by the development of transformer-based models like BERT, RoBERTa, GPT-3, and T5. These models have achieved or surpassed human-level performance on several tasks within GLUE and have made substantial progress on SuperGLUE tasks.
- **State of the Art**: The current state-of-the-art models for these benchmarks often involve large pre-trained transformer models that have been further fine-tuned on the specific tasks. Innovations in model architecture, training procedures, and data augmentation have also contributed to performance improvements.

### 2.5.15 Challenges and Limitations

- **Task Saturation**: Some tasks in GLUE have become saturated, meaning that AI models perform as well as or better than humans, leading to the creation of SuperGLUE.
- **Complexity and Diversity**: Despite these advancements, tasks in SuperGLUE remain challenging, highlighting areas where AI still struggles, such as with nuanced inference, complex reasoning, and deep commonsense understanding.

- **Beyond Benchmarks:** Achieving high scores on these benchmarks does not necessarily equate to true language understanding. There is ongoing debate about how well these tasks represent the full complexity of human language and whether models that excel on these benchmarks genuinely understand language or are exploiting statistical patterns in the data.

### 2.5.16 Conclusion

GLUE and SuperGLUE benchmarks have played a pivotal role in advancing NLU research, pushing the development of models that can tackle increasingly complex language understanding tasks. While remarkable progress has been made, these benchmarks also highlight the limitations of current models and the need for continued

⇒

## 2.6 Artificial General Intelligence (AGI) and Understanding

*Speculate on how AGI would fare in the "understanding" challenges*

Artificial General Intelligence (AGI) represents a hypothetical future stage of artificial intelligence capable of understanding, learning, and applying knowledge across a broad range of tasks and domains, essentially matching or surpassing human cognitive abilities. Given this definition, let's speculate on how AGI might fare in the understanding challenges based on the benchmarks and criteria previously discussed:

### 2.6.1 Linguistic and Semantic Understanding

AGI would likely excel in linguistic and semantic understanding, not only matching patterns in data but also grasping the deeper meanings, cultural nuances, and contextual subtleties of language. Its performance in benchmarks like GLUE, SuperGLUE, and the Winograd Schema Challenge would be at or near human levels, demonstrating an ability to interpret ambiguity, sarcasm, metaphor, and idiomatic expressions with a high degree of accuracy.

### 2.6.2 Commonsense and Reasoning

With its broader understanding and learning capabilities, AGI would bring commonsense reasoning to a new level, effectively applying general knowledge about the world to solve problems and make inferences. In challenges like CommonsenseQA and PIQA, AGI would likely show a nuanced understanding of physical interactions, social norms, and everyday situations, leveraging its integrated knowledge base to navigate questions that require extrapolation beyond specific facts.

### 2.6.3 Ethical and Moral Reasoning

AGI's performance in ethical and moral reasoning would be groundbreaking. Unlike current AI, which can simulate ethical decision-making based on data patterns, AGI would be capable of understanding the underlying principles of ethical theories and applying them to complex scenarios. Through benchmarks like the ETHICS dataset, AGI could demonstrate a sophisticated grasp

of moral dilemmas, possibly offering solutions that balance various ethical considerations and reflect a deep understanding of human values.

#### **2.6.4 Emotional Intelligence**

AGI would likely show remarkable abilities in emotional intelligence, understanding and responding to emotional cues in a way that feels authentic and empathetic. It could accurately identify emotions from text, voice, facial expressions, and other modalities, using this information to inform its interactions. In benchmarks like EmoContext and GoEmotions, AGI's performance would mirror human sensitivity and adaptability, engaging in emotionally nuanced conversations and relationships.

#### **2.6.5 Creativity and Novel Problem-Solving**

One of the most intriguing prospects of AGI is its potential for creativity and innovation. AGI could generate original ideas, artworks, theories, and solutions to novel problems, drawing on a vast, integrated knowledge base and an ability to think abstractly. In creative tasks and adversarial creativity challenges, AGI would likely produce work that is not only novel but also meaningful and insightful, pushing the boundaries of what's considered possible in artificial creativity.

#### **2.6.6 Continuous Learning and Adaptation**

AGI's ability to learn continuously and adapt to new information would fundamentally change the landscape of AI. It would not be limited by static datasets or predefined knowledge; instead, AGI would evolve its understanding through ongoing learning, similar to humans. This capacity would enable AGI to stay current with new developments, adapt its understanding based on feedback, and refine its responses over time, showcasing an unprecedented level of cognitive flexibility and growth.

#### **2.6.7 Speculative Considerations**

While this speculation paints a promising picture of AGI's capabilities in understanding, it's essential to recognize the hypothetical nature of AGI and the complex ethical, societal, and technical challenges its development would entail. Achieving AGI that truly understands in a human-like manner involves not only technological advancements but also deep philosophical inquiry into the nature of consciousness, ethics, and the essence of understanding itself.

Moreover, the development of AGI raises important ethical questions about autonomy, agency, and the implications for society, work, and human identity. As such, the journey toward AGI is not just a scientific endeavor but a profound exploration of what it means to be intelligent and understanding beings in a shared world.

## 2.7 Understanding and Explanation: What's the difference?

*Discuss the Similarities and Differences between Understanding and Explanation. Summarize in a table*

Understanding and explanation are closely related concepts but they serve different functions in the process of cognition and communication. Here's a summary of their similarities and differences:

Aspect	Understanding	Explanation
Definition	The ability to grasp the meaning, significance, or nature of something.	The act of making something clear or understandable.
Purpose	To comprehend the how and why of phenomena, concepts, or information.	To provide reasons, justifications, or clarify how and why something occurs.
Process	Internal cognitive process involving interpretation, contextualization, and integration of information.	External communication intended to make information understandable to others.
Direction	Inward (subjective experience and cognitive processing).	Outward (aimed at another's understanding or clarification).
Outcome	Personal insight, comprehension, or grasp of concepts.	Increased understanding, clarity, or awareness in others.
Dependency	Can occur independently, as it is a personal process.	Requires an audience or a context; it is relational.
Examples	Understanding the principles of physics.	Explaining how gravity works.
Nature	Often silent and implicit.	Typically verbal, visual, or written and explicit.
Requirement	Requires cognitive and sometimes emotional processing.	Requires ability to articulate or convey information effectively.
Evaluation	Subjective and can be difficult to measure externally.	Can be assessed based on clarity, coherence, and reception by the audience.

While understanding and explanation are interrelated—effective explanations can facilitate understanding, and deep understanding can improve the quality of explanations—they cater to different aspects of cognitive and communicative processes.

⇔

## 2.8 Desperately Seeking Understanding

Here we shift gears to exploring the discussion of *Understanding* in general.

## 2.9 Looking for books with understanding in their title

*User: Identify major books discussing "understanding" that contain the word "understanding" in their titles.*

Here are some notable books that discuss various aspects of understanding and contain the word understanding in their titles, spanning different fields such as philosophy, psychology, education, and science:

### 1. **Understanding Media: The Extensions of Man by Marshall McLuhan**

- This book is a seminal work in media studies, exploring how media influences human perception and understanding of the world.

### 2. **Understanding Comics: The Invisible Art by Scott McCloud**

- A comic book that itself explains the history, vocabulary, and methods behind comics and how they are understood by the human mind.

### 3. **Understanding Power: The Indispensable Chomsky by Noam Chomsky**

- This collection of edited transcripts from seminars and discussions reveals Chomsky's perspectives on power and politics and how they shape our understanding of the world.

### 4. **Understanding Phenomenology by Michael Hammond, Jane Howarth, and Russell Keat**

- This book provides an introduction to phenomenology, a philosophical movement that emphasizes the structures of experience and consciousness.

### 5. **Understanding Syntax by Maggie Tallerman**

- Aimed at students and researchers in linguistics, this book explains the fundamental concepts of sentence structure and the understanding of syntax.

### 6. **Understanding Motivation and Emotion by Johnmarshall Reeve**

- This psychology book explores the processes behind human motivation and emotions, offering insights into how they can be understood and influenced.

### 7. **Understanding Machine Learning: From Theory to Algorithms by Shai Shalev-Shwartz and Shai Ben-David**

- This book provides an introduction to the field of machine learning, presenting both theoretical concepts and practical algorithms.

### 8. **Understanding Digital Culture by Vincent Miller**

- This work examines the impact of digital technology on various aspects of culture and how we understand our contemporary world.

**9. Understanding International Relations by Chris Brown and Kirsten Ainley**

- An introductory text that explains the major principles, theories, and methodologies within the field of international relations.

**10. Understanding Earth by John Grotzinger and Thomas H. Jordan**

- A textbook on earth sciences, offering insights into the physical and chemical principles that govern the Earth and its environment.

These books cover a range of disciplines, reflecting the broad applicability and importance of understanding as a concept in various fields of study.

*Observation:* Chat complied with the request, but not with its spirit of discussing the cognitive / mental process of understanding.

⇒

*Usere:* thank you for your response; while technically correct that they have the word understanding in their title, I am really looking for scholarly works that area about understanding and contain the word understanding in their title. A great example of that is John Locke's "An Essay Concerning Human Understanding". So repeat the task again, but now focusing on scholarly work that focus on the mental cognitive aspect as described by Locke

Certainly! Here are some scholarly works that focus specifically on the cognitive and mental aspects of understanding, with understanding in their titles:

**1. An Essay Concerning Human Understanding by John Locke**

- This foundational philosophical work explores the nature of human knowledge and understanding, arguing that all knowledge derives from experience.

**2. Understanding Understanding: Essays on Cybernetics and Cognition by Heinz von Foerster**

- This collection delves into the concepts of cybernetics, cognition, and how understanding emerges within these contexts.

**3. Understanding Phenomenal Consciousness by William S. Robinson**

- This book addresses the challenges of understanding consciousness and the qualitative aspects of sensory experiences.

**4. Structures of the Mind: Essays on Language, Music, and Cognition in Honor of Ray Jackendoff**

- While not authored by Ray Jackendoff, this collection includes essays that reflect on his work related to the cognitive structure of understanding in various domains.

**5. Understanding Others: Peoples, Animals, Pasts by Rowlands Mark**

- A philosophical examination of how we understand other minds, including those of animals and individuals from different historical contexts.

**6. The Understanding by Design Guide to Creating High-Quality Units by Grant Wiggins and Jay McTighe**

- This educational work discusses the ‘backward design’ framework, emphasizing understanding in learning and teaching strategies, albeit with an educational slant.

**7. Making it Explicit: Reasoning, Representing, and Discursive Commitment by Robert B. Brandom**

- This philosophical text addresses the nature of understanding through the lens of language, reasoning, and social practices.

**8. Understanding Emotions by Keith Oatley, Dacher Keltner, and Jennifer M. Jenkins**

- A scholarly work focusing on the cognitive aspects of understanding emotions, their causes, and effects.

**9. Sources of Power: How People Make Decisions by Gary Klein**

- While not strictly in the title, this book is a fundamental work in understanding how people process information and make decisions under pressure.

**10. Mind and World by John McDowell**

- This book deals with the fundamental structure of understanding and how our mind relates to the world, including the nature of perception and conceptualization.

These works span a range of disciplines, including philosophy, psychology, education, and cognitive science, reflecting the multifaceted nature of understanding as a mental and cognitive process.



*User: those are nice, but you provide many titles that don't have the word "understanding" explicitly in their title. Repeat the task again with additional titles that explicitly include the words understanding in their title*

I apologize for the oversight. Here's a refined list focusing strictly on scholarly works that have understanding explicitly in their title, and which focus on the mental or cognitive aspects similar to the theme explored by John Locke:

**1. An Essay Concerning Human Understanding by John Locke**

- This seminal philosophical work investigates the origin, extent, and certainty of human knowledge, arguing that all knowledge is derived from sensory experience and internal reflection.

**2. Understanding Understanding by Richard Mason**

- A philosophical exploration into what it means to understand something, discussing the conditions under which understanding occurs and its significance in human cognition.

**3. Models of Understanding Text edited by Britton and Graesser**

- This collection addresses different psychological models and theories about how people understand written texts, relating to cognitive processes and comprehension.
4. **Understanding Emotions by Keith Oatley, Dacher Keltner, and Jennifer M. Jenkins**
    - This book delves into the cognitive, social, and physiological aspects of emotions and how they contribute to our understanding of the world and ourselves.
  5. **Understanding Other Minds: Perspectives from Developmental Social Neuroscience by Simon Baron-Cohen, Helen Tager-Flusberg, and Michael Lombardo**
    - A comprehensive text on the cognitive processes underlying our ability to understand and interpret the mental states of others, known as 'Theory of Mind'.
  6. **Understanding Consciousness by Max Velmans**
    - A significant work that combines cognitive science and psychology to explore the nature of consciousness and how it can be understood from a scientific perspective.
  7. **Understanding Syntax by Maggie Tallerman**
    - While primarily linguistic, this book delves into cognitive aspects of how humans understand sentence structure and linguistic patterns.
  8. **Philosophy and the Sciences of Exercise, Health and Sport: Critical Perspectives on Research Methods edited by Scott Kretchmar and others**
    - This collection, while not exclusively about mental processes, includes discussions on understanding human behavior and cognition in the context of physical activity.
  9. **Understanding Computers and Cognition: A New Foundation for Design by Terry Winograd and Fernando Flores**
    - Although it focuses on computers, this book delves deeply into how understanding and cognition can inform the design of information systems.
  10. **Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference by Zoltan Dienes**
    - This book provides an introduction to the philosophy of science through the lens of psychology, focusing on how understanding develops through scientific methods.

These titles are more aligned with the specific request for works that include understanding in their titles and focus on cognitive, psychological, or philosophical aspects of understanding.



*Observation:* Chat came up with some real gems with this reply! I didn't check every title for potential hallucinations; the ones I did check were definitely legit and looked to be of high quality. Also, totally new to me. I am aware of a number of major titles, but clearly not all.

⇔

*User: that looks good. Keep going*

Sure, here are more scholarly works specifically focusing on the cognitive and mental aspects of understanding, where understanding is explicitly included in the title:

1. **Understanding Understanding: Essays on Cybernetics and Cognition by Heinz von Foerster**
  - This collection discusses the foundations of knowledge and understanding from a cybernetic perspective, touching upon how systems (including humans) perceive and understand their environments.
2. **Models of Understanding Text edited by Bruce K. Britton and Arthur C. Graesser**
  - This book compiles various essays on how people understand texts, incorporating perspectives from psychology, education, and linguistics to offer a comprehensive view on text comprehension.
3. **Understanding Minds and Brains edited by Paul Thagard**
  - A collection of essays that explore different approaches to understanding how minds and brains work, incorporating insights from cognitive science and philosophy.
4. **Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience by Simon Baron-Cohen, Helen Tager-Flusberg, and Michael Lombardo**
  - This work delves into how we come to understand the minds of others, with a focus on the development of theory of mind and its implications for cognitive neuroscience.
5. **Understanding Consciousness by Max Velmans**
  - Velmans explores the complex nature of consciousness from a psychological standpoint, questioning what it means to understand conscious experiences.
6. **Understanding Emotions by Keith Oatley, Dacher Keltner, and Jennifer M. Jenkins**
  - This book provides a detailed exploration of the cognitive underpinnings of emotions and how they influence our understanding of the world and ourselves.
7. **Understanding Syntax by Maggie Tallerman**
  - While primarily a linguistic text, this book delves into the cognitive processes involved in understanding the structure of language.

**8. Understanding the Representational Mind: A Prerequisite for Inter-subjective Communication by Josef Perner**

- Perner's work examines the cognitive foundations necessary for understanding the mental states of oneself and others, crucial for effective communication and social interaction.

**9. Understanding and Preventing Violence: The Psychology of Human Destructiveness by Leila B. Leonard**

- This scholarly work looks into the cognitive and psychological factors leading to violence, aiming to deepen the understanding of violent behavior and its prevention.

**10. Understanding Sleep and Dreaming by William H. Moorcroft**

- An exploration of the cognitive and psychological processes involved in sleep and dreaming, providing insights into how and why we dream and what it reveals about our understanding of the human mind.

These titles provide a deeper dive into the concept of understanding from various scholarly perspectives, focusing on how it is conceptualized and studied within the fields of cognitive science, psychology, education, and neuroscience.



*User: what are books similar to John Locke's "An Essay Concerning Human Understanding"*

Books similar to John Locke's *An Essay Concerning Human Understanding* typically delve into the nature of human cognition, knowledge, and the philosophical underpinnings of the mind. Here are some classic and influential works in the realm of philosophy and epistemology that share similar themes:

**1. Critique of Pure Reason by Immanuel Kant**

- This seminal work in philosophy attempts to understand the limits and scope of human knowledge and reason. Kant explores how we perceive the world and the distinctions between appearances and the nature of things themselves.

**2. Meditations on First Philosophy by René Descartes**

- In this work, Descartes examines the foundational aspects of human knowledge and understanding, famously concluding with the assertion *I think, therefore I am*. He explores doubt, existence, and the nature of reality.

**3. The Problems of Philosophy by Bertrand Russell**

- Russell addresses the principles of epistemology and the questions surrounding human knowledge, its acquisition, and its limitations, providing an accessible introduction to philosophical inquiry.

**4. An Enquiry Concerning Human Understanding by David Hume**

- Hume critiques the notion of human understanding explored by earlier philosophers, especially focusing on concepts of causation, personal identity, and the limits of human knowledge.

## 5. **The Phenomenology of Spirit by G.W.F. Hegel**

- Hegel's complex work deals with themes of knowledge, consciousness, and the nature of reality. He explores the development of self-consciousness and the dialectical process of understanding.

## 6. **Tractatus Logico-Philosophicus by Ludwig Wittgenstein**

- While more focused on logic and language, Wittgenstein's work touches upon how language frames our understanding of the world and discusses the limits of what can be known and expressed.

## 7. **Philosophical Investigations by Ludwig Wittgenstein**

- In this later work, Wittgenstein moves away from the ideals of pure logic to explore how language, practice, and forms of life shape our understanding of the world.

## 8. **Being and Time by Martin Heidegger**

- Heidegger delves into the nature of being and the human condition, exploring concepts such as existence, time, and consciousness, which contribute to our understanding of ourselves and the world.

## 9. **Knowledge and Its Limits by Timothy Williamson**

- This is a more contemporary work that discusses the nature of knowledge, its scope, and its limitations, offering a detailed theory called epistemicism.

## 10. **The Structure of Scientific Revolutions by Thomas S. Kuhn**

- While focused on the history and philosophy of science, Kuhn's work provides profound insights into how scientific understanding evolves and is influenced by paradigms.

These works, like Locke's essay, are foundational texts in the philosophical exploration of knowledge, understanding, and the workings of the human mind. They offer varying perspectives and critiques that have shaped Western philosophical thought.

⇒

## **2.10 Discussing of Locke's *An Essay Concerning Human Understanding***

*User: provide a very granular discussion of John Locke's "An Essay Concerning Human Understanding"; summary by listing the main points.*

An Essay Concerning Human Understanding by John Locke is a foundational philosophical text that explores the nature of human knowledge and understanding. Here's a granular summary of its main points:

### **2.10.1 Book I: Of Innate Notions**

1. **Rejection of Innate Principles:** Locke argues against the notion that there are innate principles or ideas present in the mind from birth. He challenges the existence of universally accepted truths and suggests that, had such innate ideas existed, all human beings would share them.

2. **The Mind as a Blank Slate (Tabula Rasa):** He introduces the concept that the mind at birth is a white paper, void of all characters, with experience writing upon it. This contrasts sharply with the prevailing notion of innate ideas or knowledge.

### 2.10.2 Book II: Of Ideas

1. **Origin of Ideas:** Locke states that all ideas come from experience, divided into two categories: sensation (external sensory experience) and reflection (internal mental processes).
2. **Simple and Complex Ideas:** He distinguishes between simple ideas, which are elemental and unanalyzable, and complex ideas, which are constructed from simple ones through combination, comparison, and abstraction.
3. **Quality and Substance:** He explores the concept of primary qualities (inherent in objects) and secondary qualities (perceived through senses but not actually in objects), influencing the distinction between subjective and objective reality.
4. **Language and Meaning:** Locke discusses the role of language in human understanding, noting its necessity for communication but also its limitations and the potential for misunderstanding.

### 2.10.3 Book III: Of Words

1. **The Signification of Words:** He delves into the nature of language, the use of words as symbols for ideas, and the problems of ambiguity and misinterpretation that arise from language.
2. **Abstract General Ideas:** Locke examines the construction of general and abstract terms, arguing that they are necessary for higher-level reasoning and communication but are rooted in particular instances.

### 2.10.4 Book IV: Of Knowledge and Probability

1. **Nature and Extent of Human Knowledge:** Locke outlines the limits of human knowledge, distinguishing between intuitive, demonstrative, and sensitive knowledge, and admitting that much of our reasoning relies on probability rather than certainty.
2. **Degrees of Assent:** He describes the varying degrees of assent people give to different propositions and the influence of reason, revelation, faith, and enthusiasm on belief and opinion.
3. **The Limits of Knowledge:** Locke asserts that there are bounds to what humans can understand, influenced by the limitations of our senses, the complexity of the material world, and the finite capacity of our minds.
4. **Faith and Reason:** He distinguishes between knowledge and belief, stating that faith should not contradict reason, and that revelation can be a legitimate source of knowledge if it aligns with reasoned understanding.

⇒

### 3 Brass Tacks: Getting down to business

The next step in my expedition, was to figure out what *Chat itself* says it's doing. This is not straightforward as obviously OpenAI would like to keep its Secret Sauce to itself. Also, despite experimenting with Chat for over a year, I have lots more learning to do. It feels like there are endless variations and approaches possible for interacting (prompt crafting and interpretation) with Chat. It seems like I'm getting better with time, and of course Chat itself is morphing and growing.

#### 3.1 Round One

The idea here is for me to create a pre-prompt prompt (or *metaprompt* if you will). The metaprompt is aimed at prompting Chat to suggest a prompt optimized for working with GPTs.

*Metaprompt: Create an effective GPT prompt that explores definitions and relationships based on the following thought: "GPT Understanding; How do you as a GPT understand prompts". The prompt should ask for granular step by step consideration of the ideas, items, and relationships under consideration, and conclude with a table summarizing key elements and relationships*

Creating an effective GPT prompt to explore definitions, relationships, and understanding can be crucial for obtaining precise and useful information. Here's an example of how you could structure your prompt:

---

#### **Prompt for GPT:**

Title: Understanding GPT's Interpretation of Prompts

Introduction:

Begin by explaining what prompts are and their role in interacting with AI models like GPT. Describe how prompts serve as the primary method of communication between humans and GPT, directing the AI's focus and responses.

Definitions:

1. Define 'GPT' and 'Prompt' in your own terms.
2. Explain the concept of 'Understanding' in the context of AI, particularly how GPT interprets text.

Process of Interpretation:

1. Describe the step-by-step process you, as a GPT, undergo from receiving a prompt to generating a response. Include initial interpretation, context analysis, and relevant information retrieval.
2. Detail how you assess the relationships between different concepts within a prompt.

Factors Influencing Responses:

1. Identify and explain the factors that influence how you interpret prompts and generate responses, including language structure, clarity of the prompt, and previous information provided in the conversation.

Case Study:

1. Provide a detailed example of how you would break down and interpret a complex prompt. Use a hypothetical prompt related to environmental science, breaking down terms, implied questions, and
2. expected types of responses. Explain how different interpretations of the prompt could lead to different responses.

Summary Table:

Create a table summarizing the key elements discussed:

- The term (GPT, Prompt, Understanding)
- Definition
- Relationship to other terms
- Impact on GPT's response process

Conclusion:

Conclude by summarizing how understanding the structure and elements of a prompt can lead to more effective and accurate GPT responses. Reflect on the importance of clear communication between humans and AI.”

---

This structured prompt encourages a comprehensive exploration of the concept of understanding within the context of AI interactions, prompting a detailed explanation and analysis that should culminate in a clearer understanding of how GPT processes and responds to user inputs.

⇒

The next step is to enter the above, pretty detailed prompt, and see what Chat actually does with it. So I did that and its response was:

⇒

Based on the structured prompt you provided, here's how a GPT model might address each section:

## **Title: Understanding GPT's Interpretation of Prompts**

### **Introduction:**

Prompts are instructions or questions directed towards an AI model, providing a framework for the required response. In the context of AI, like GPT, prompts act as the interface for interaction, guiding the AI's focus, understanding, and subsequent responses. This interaction is foundational for achieving meaningful and relevant AI output.

### **Definitions:**

1. **GPT:** Generative Pre-trained Transformer, an AI model designed to generate text based on the input it receives. It learns from a vast dataset to understand and replicate human-like text
2. **Prompt:** A text input given to AI models like GPT to elicit a specific type of response or information.

3. **Understanding in AI:** The process by which AI interprets, processes, and assigns meaning to the input text, based on trained patterns and context

**Process of Interpretation:**

1. **Initial Interpretation:** GPT starts by parsing the prompt to identify keywords, phrases, and the overall topic or question.
2. **Context Analysis:** The model evaluates the prompt within the larger context of the conversation or based on its training data.
3. **Information Retrieval:** GPT pulls relevant information and structures a response based on learned patterns and the specifics of the prompt.
4. **Assessing Relationships:** The AI identifies and considers the relationships between different concepts mentioned in the prompt to ensure coherent and contextually appropriate responses.

**Factors Influencing Responses:**

1. **Language Structure:** The clarity, grammar, and complexity of the prompt influence how the AI interprets and responds.
2. **Clarity of the Prompt:** Clear, direct prompts lead to more accurate responses, while vague prompts may result in generalized answers.
3. **Previous Information:** If the prompt is part of an ongoing conversation, previous exchanges impact the AI’s response.

**Case Study:**

1. **Complex Prompt Breakdown:** If presented with an environmental science prompt, GPT would dissect the terms (e.g., ‘biodiversity’, ‘climate change’), interpret implied questions, and craft responses that align with the expected information (definitions, impacts, solutions).
2. **Different Interpretations:** Variations in the prompt’s wording can lead to different angles of response, emphasizing certain aspects over others based on perceived importance or relevance.

Table 3: **Summary Table:**

Term	Definition	Relationship to other Terms	Impact on GPT’s Response Process
GPT	AI model generating human-like text	Uses prompts for guidance	Determines response structure and content
Prompt	Text input to elicit specific responses from GPT	Directs GPT’s focus	Influences clarity and specificity of GPT’s output
Understanding	The process of interpreting and processing input in AI	Basis for generating responses	Affects accuracy and relevance of responses

**Conclusion:** Understanding the nuances of how prompts are structured and interpreted by GPT can significantly enhance the effectiveness and accuracy of the AI's responses. Clear communication, contextual awareness, and a structured approach to prompting ensure more meaningful interactions between humans and AI, highlighting the importance of precision in language and intent.

⇔



## 3.2 Round Two

*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 ideas 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.

*User:* Thank you, let's be more specific and discuss how you CHAT (GPT4) conduct understanding of prompts. Explain the step by step way on how YOU understand prompts

Certainly! Here's a step-by-step explanation of how I, as a GPT-4 model, understand and process prompts:

### 3.2.1 Initial Reception:

- **Input Parsing:** When I receive a prompt, the first step is parsing the input text. This means I break down the text into manageable pieces, such as sentences and words, to understand its basic structure.

### 3.2.2 Preprocessing:

- **Tokenization:** I convert the text into tokens, which are essentially pieces of words or entire words that help me understand the text in the way humans use language.
- **Normalization:** I standardize the text by converting it to lowercase, removing punctuation, and performing other clean-up tasks to reduce variability in the input.

### 3.2.3 Understanding Context:

- **Immediate Context:** If the prompt is part of a conversation, I look at the immediate context, which includes previous interactions in the session. This helps me maintain coherence and relevance in the conversation.
- **Global Understanding:** I use the knowledge trained into me from a vast array of texts to understand the general context of the topic or question.

### 3.2.4 Semantic Analysis:

- **Identifying Key Components:** I identify key components of the prompt, such as the main subjects, verbs, objects, and any specific demands or questions.
- **Understanding Intent:** I infer the intent behind the prompt — whether it's asking for information, requesting an action, or seeking an opinion.

### 3.2.5 Response Formulation:

- **Information Retrieval:** Based on the prompt's content and context, I retrieve relevant information from my training data.
- **Answer Structuring:** I organize the retrieved information logically, ensuring it aligns with the prompt's intent and format.

### 3.2.6 Language Generation:

- **Drafting Response:** I generate a draft response based on the structured information and intended message.
- **Refinement:** I refine the response for coherence, grammar, politeness, and alignment with conversational norms.

### 3.2.7 Response Evaluation:

- **Self-Evaluation:** Before presenting the response, I evaluate it against the original prompt to ensure relevance and accuracy.
- **Adjustment:** If necessary, I make adjustments to better align the response with the prompt's intended meaning and context.

### 3.2.8 Output Delivery:

- **Final Response:** Once I am satisfied with the relevance and coherence of the response, I present it as the output.

This process is highly iterative and involves complex algorithms and neural network mechanisms. It's important to note that while I strive for accuracy, my responses are based on patterns and information in the data I've been trained on, and I don't access or retrieve information from the internet or external databases in real-time.

*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.

⇐

## 4 Assessing Chat's Understanding of Understanding

OK, so this was an interesting exercise. I didn't know what I would get. I created follow-up questions after reading Chat's responses. My high level observations are;

1. The responses were useful! Some of them were simplistic, but then again, I've studied and explored these areas before so I was aware of a fair amount of what Chat responded with. Others that have less acquaintance with the subject matter would find the responses great places to start with. Some of the responses were interestingly and usefully detailed
2. Lots of caveats all over the place. Chat doesn't want to come across as an ultimate arbiter or authority.
3. Overly pedantic about ethical, diversity and inclusion reminders. This *appears too forced*.
4. I haven't checked all the references provided. I can't find some of them using standard Google Searches. It seems like some of the references are 'close', meaning they refer to an authoritative, well known, author that has books or papers on adjacent subjects with titles that use some of the same substantive anchor words in their titles.
5. I did find some *gems* in the responses that I wasn't aware of and I look forward to reviewing and reading those in depth.
6. Responses to requests for detailed description of technical approaches in some AI related areas are very substantive and exposed new information (to me). Many serendipitous possibilities here.
7. Temporal Awareness: Puzzling. When discussing 'current state of the art', or recent timing, Chat seemed to use references to past, current, and future events in somewhat inconsistent ways. Possibly, some of its training used material that refers to future events, where the future time has, in reality, already passed. This is a challenge for training in general, and especially training that may take months per complete training and release cycle. I imagine at some point, all response will need to go through additional scrubbing before shown on the User's screen.

⇒

## 5 Closing Thoughts

This was an interesting experience. I learned a variety of things across disciplines and the art of prompting. One learns how to be more judicious and careful with one's words. What and how you phrase requests / prompts really make a difference.

Clearly, Chat enables an incomprehensible vast range of possible directions to pursue. I started with the goal of exploring David Deutsch's comment,

"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.**"

Lets look at two parts of this quote:

- a) "Understanding is one of the higher functions of the human mind and brain, and a unique one" and,
- b) "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"

(a) refers to understanding as a function of brain and mind. We do have a reasonable understanding of the material construction of brains. Its made up of molecules organized at low level components that make up larger functional units, and then even larger components (e.g neurons, synapses, neurotransmitters, amygdala, frontal cortex . . . ). There are lots and lots of components. We probably haven't identified all of them, but we're pretty close. While we can distinguish certain types of neurons that have particular functions and construction (Table 4 lists some of these neural cells). We really don't have a solid grasp of what mind is! There's an enormous amount of writing, speculation, perceptions, experiences all dealing with minds. Discussing minds intelligently can easily take a lifetime of work.

(b) refers to two principle interests:

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

Some of the challenges ahead are then making progress with specifics associated with (a) and (b). This progress needs to be actual progress, in the sense the brilliant physicist Bridgeman<sup>30</sup> associated with operational methods and operational philosophy, for example that described in his Quo Vadis article<sup>31</sup>. 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.

<sup>30</sup>[https://en.wikipedia.org/wiki/Percy\\_Williams\\_Bridgman](https://en.wikipedia.org/wiki/Percy_Williams_Bridgman)

<sup>31</sup>Bridgman, P. W. "Quo Vadis." Daedalus 87, no. 1 (1958), pp. 85-93 <http://www.jstor.org/stable/20026428>

The contrast between theoretical or abstract and operational approaches is that theoretical or abstract ones remain in the realm of ideas and concepts, while those that are operational are worked with or tested on actual existing (or physically constructable) devices. Success of demonstration of the understanding of concepts, ideas, or theories, are deemed actualized when a test or demonstration is conducted on a physical device. When dealing with abstraction in computations or for example mathematical theories; operational success is realized when the idea, theory, or concept is demonstrated via specific calculations with actual numbers. Mathematics does allow for other forms of operational demonstrations; those need to ultimately be tested with specific cases.

Table 4: **Some Neural Cell Types**

Neuron Type	Key Function	Distinguishing Features
Pyramidal Neurons	Cognitive functioning, learning, memory	Pyramid-shaped cell body, large dendritic tree, long axon
Interneurons	Modulation, inhibition, synchronization of neural activity	Usually small, highly connected, mostly use GABA
Purkinje Cells	Motor control, coordination	Large, elaborate dendritic arbor, inhibitory
Granule Cells	Memory processing, motor coordination	Small, high density of dendritic spines
Motor Neurons	Activation of muscles or glands	Long axons, connect CNS to muscles/glands
Sensory Neurons	Transmission of sensory information	Long dendrites and axons, part of the PNS
Oligodendrocytes	Support and insulation of CNS neurons	Form myelin sheath in CNS
Schwann Cells	Support and insulation of PNS neurons	Form myelin sheath in PNS, aid in nerve regeneration

To cast the challenge of (a), (b.1 and )(b.2) as *Bridgemanian operational challenges* we would need to localize where and how *understanding* occurs in the brain and develop an operational approach for discussing mind (and then locating or identifying testable aspects of *understanding* in minds). Now, since we're also discussing *intelligence* in machines, we would need to develop highly detailed operational descriptions of components and algorithms that are involved in execution of processes of understanding. Challenge (b.2), is more substantially difficult. Wanting one (an explanation) in humans is an intentional psychological behavioral state. I'm very likely not using the correct words here. The question of machines *wanting explanations* is a complex arena. Sure, one could code strings of words and even conditional or randomly generated set to goals to execute; how these relate to what human wants and their causes is rather complex. I am not saying this cannot be done; it has not been done yet in the sense that Deutsch referred to.

To recap, at the start of this exercise, I was curious and wanted to get a sense of whether the statement "**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.**" holds. I won't declare a final definite judgment yet. It does seem that the "capable of" part is at hand or almost at hand. The "wanting one" . . . hmm, how would we be able to tell?

In discussing Artificial Intelligence one often runs into discussions of the *Turing Test*, originally called the imitation game by Alan Turing. Wikipedia summarizes "a test of a machine's ability to exhibit intelligent behaviour equivalent to, or indistinguishable from, that of a human"<sup>32</sup>. In my opinion, with respect to the specific implementation as Turing described it, we're converging on declaring it as "an interesting historical benchmark that has been satisfied by sufficiently Large Language Models". To me, it is time to retire it, and establish a more challenging intelligence test. Maybe a *Newton Test* that would be judged by the ability of machine intelligence, or perhaps AGI, to describe a very challenging unexplained (as of yet) fundamental science real world, and then solve it in an elegant way.

This elegant way would entail actual repeatable experiments that would test the various components used in 'solving' the problem, and would withstand all challenges. Something akin to Newton's establishing that white light (daylight) is actually made of multiple light 'components' (light of different colors / frequencies), that could and were resolvable by the use of a prism. That was elegant, advanced the science of optics, and absolutely repeatable.

#### **What's Next? Queuing up future Chats.**

In the grand EP tradition, one must disclose future journeys and expeditions of the United Federation of Galaxies' Pizza Delivery Scooter *Enterprise*. Recalling its continuing mission 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, the following expeditions are planned:

- More Chats with Chat!
- Improving the understanding of *understanding*, specifically with respect to items (b.1) - Capability of understanding an explanation, and (b.2) - Wanting one (an explanation), discussed earlier.
- IJ Good's UltraIntelligent Machines
- Exploring the notion of self, self-awareness, self-reference in intelligent machines.
- Introducing and exploring the *Understand -Decide - Act* (UDA) Arc. Moving from John Boyd's *OODA Loop*<sup>33</sup> to the *UDA ARC*. Exploring GPT internal UDA Arcs with Chat would be rather fascinating.

---

<sup>32</sup>[https://en.wikipedia.org/wiki/Turing\\_test](https://en.wikipedia.org/wiki/Turing_test)

<sup>33</sup>[https://en.wikipedia.org/wiki/OODA\\_loop](https://en.wikipedia.org/wiki/OODA_loop)

My main summary:

- This paper presented a conversational chat with ChatGPT.
- At a personal level, I can say it was quite enjoyable and I learned a lot.
- Even at this early stage in development of Large Language Models and GPTs, the results are impressive and useful.
- One can develop strategies, like the MetaPrompt (discussed earlier) to substantially improve Chat responses.
- While there hallucinations are possible, and must be guarded against, Chat can and does provide some *raw gems*. Careful users can benefit substantially from these.
- Depending on the way one uses Chat, its somewhere between a starting smart intern and a Postdoc. It does provide really good Python and LaTeX code snippets (my use cases). On multiple occasions the code provide worked without modification (on my MacOS system).
- I will continue my *Chat with ChatGPT* explorations. I look forward to improving my strategy and approach with an eye to *operational applications* in the spirit of Bridgeman.
- If you haven't done so yet, I recommend you explore Chat (or another comparable LLM)
- There's a lot more to be done. In addition to improving the understanding of *understanding*, I will be conducting Chats with ChatGPT on specific topics described in the next section.

⇒

## 6 Dedication and Acknowledgments

**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 Large Language Models (LLMs), in particular Open AI's ChatGPT and DALLE The work itself is drafted using  $\LaTeX$ , TeXstudio 4.7.2 Using Qt Version 6.6.0. ChatGPT output obtained in Markdown format was converted 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.



FIN  
⇌  
FIN