

Definable Units of "Intelligence" for Evaluating AI Performance

Object Relationship Spaces for AI-ML: A Framework for
Clearly Defined, STEM-Compatible, Project-Level, Functional Units of "Intelligence"
For AI Design, Analysis, Performance, Architecture, and Operating Systems
G.G.Ashbrook, 2023.03/04

Abstract

There is a need for the use of well defined performance frameworks to describe the goals and skills/abilities of systems including AI.

The overall agenda here is to move toward clearer communication and better definitions, including the pragmatic utilization of universal intersecting/interlocking areas.

This proposed object-relationship-space framework can be used for guiding project-specific system design, goal-setting, discussion, testing, analysis, reporting, regulation, documentation, etc.

For more detail on what is meant by 'design': to manage and enable smaller or larger scale AI projects coordinating required abilities across internal and external components, including "symbolic" logistics and "sub-symbolic" training (including for AI-self-management), and whole operating-systems for AI.

AI must be able to handle "objects" in the following interlocking contexts:

1. object-relationship-spaces
2. (internal/external) project-object-database (in a project-framework)
3. project-participants (in a project-framework & participation-space)

such that, 'objects' are defined as existing outside of the AI for managing the project, and that so long as the AI effectively deals with these project-objects, it does not matter how the AI handles the objects 'internally'. Example alternative methods of internal handling/processing/management include:

- symbolic vs. sub-symbolic
- single vs. multiple
- mono, 'end-to-end' vs. hybrid
- parametric, nonparametric
- explainable vs. black-box
- higher dimensional vs. lower dimensional
- calculation vs. intuitive pattern recognition
- similar-to-h.sapiens, vs. not similar to h.sapiens
- etc.

Interconnecting/Intersecting Areas:

A repeating theme, context, and agenda in this paper is to pragmatically leverage the interconnected functionality of clear definitions, STEM, projects, participation, positive values, and productivity.

To reiterate and state this as openly as possible, the context and agenda here is a project, best practice, positive-values, productivity, context.

Framework vs. Pronouns vs. Hyperbole vs. Tesler Rule (Summary)

1. Discussion with undefined terms (for example specific abilities) can loop indefinitely regardless of the abilities of AI at that time or changes over time in what AI can do: Undefined & under-defined goals, terms and definitions tell us too little about what is needed, what the system can do, and if the system can do what is needed.

2. Telser Rule Loop: where 'AI' is undefined and every new development is dismissed as "not real AI." Refusing to define 'AI' tells us too little about what is needed, what the system can do, and if the system can do what is needed.

3. According to an Object-Relationship-Space framework in a project and participation context, AI-ML technology can as of March, 2023 join *h.sapiens-humans* as a participant in projects, with specific skills/abilities to handle specific project-objects, where projects, participants, and objects in object relations spaces, are clearly and functionally defined in a STEM context. This Object-Relationship-Space framework should define what is needed, what the system can do, and that the system can do what is needed.

Part one concerns a brief overview of the framework.

Part two concerns using the framework,
e.g. so you can construct your own well defined goals and tests for abilities of AI systems.

Part three concerns a discussion of the discussion of AI,
e.g. so you can critique statements in what you read about AI.

Part four concerns goals and agendas, background concepts and principles, and future design factors.

Appendices include more examples and details.

Table of Contents: Brief

Abstract

A Narrative Introduction in Two Parts

Part 1: Framework in a Nutshell

- The Object Relationship Space Framework
- 1.1 Example Object Relationship Space List
- 1.2 Many lists in One
- 1.3 Networked-AI Components

Part 2: Using The Framework

- 2.1 Examples: AI Skills, Comparing Three Chatbots
- 2.2 Examples: AI Skills Mapped to Object Spaces
- 2.3 Adding Levels, Adding Steps, Adding Objects
- 2.4 Hybrid/cross-model skills for AI
- 2.5 Heuristic & Pseudocode for AI Management with Object Spaces
- 2.6 Modularity, Scale, and AI-Component Networks
- 2.7 Model 'Explainability' as 'Explainability, Reliability, and Security'
- 2.8 Mapping a general problems-space for AI & Mind
- 2.9 Object Relationship Based Testing

Part 3: Discussing the Discussion of AI

- 3.1 Definitions of Terms
- 3.2 Data Sources Discussing AI
- 3.3 Examining Tests for AI
- 3.4 What do we do with ChatGPT?

Part 4: Goals, Background & Future:

- 4.1 Agenda & Goals
- 4.2 Background Concepts and Principles
- 4.3 Future Design Factors

Appendices

Table of Contents: Detailed

Abstract

A Narrative Introduction in Two Parts:

Introduction Part 1. Chess in Blade Runner

Introduction Part 2. Defining AI Goals and 'Objects'

Part 1: Framework in a Nutshell -> The Object Relationship Space Framework

- 1.1 Example General Object Relationship Space List
- 1.2 Many lists in One
- 1.3 Networked-AI Components

Part 2: Using The Framework

- 2.1 Examples: AI Skills, Comparing Three Chatbots
- 2.2 Examples: AI Skills Mapped to Object Spaces
- 2.3 Adding Levels, Adding Steps, Adding Objects
- 2.4 Hybrid/cross-model skills for AI
- 2.5 Heuristic & Pseudocode for AI Management with Object Spaces
- 2.6 Modularity, Scale, and AI-Component Networks: and the need for a literal or proverbial operating system.
- 2.7 Model 'Explainability' as 'Explainability, Reliability, and Security'
- 2.8 General Mind-Space: Mapping a general problem-space for AI
- 2.9 Object Relationship Based Testing

Part 3: Discussing the Discussion of AI

(In Summary)

- 3.1 Definitions of Terms
- 3.2 Data Sources Discussing AI
- 3.3 Examining Tests for AI
- 3.4 What do we do with ChatGPT?

(In Detail)

- 3.1 Definitions of Terms
 - 3.1.1 Terminology Issues 1: The tangled Semantics of human ability.
 - 3.1.2 Terminology Issues 2: What has been defined or is not-defined?
 - 3.1.3 Terminology Issues 3: Navigate Jargon Pragmatically
 - 3.1.4 Terminology Issues 4:
Fictional frames of reference are bad:
 - 3.1.5 Terminology Issues 5:
Problematic multiple meanings of unavoidable terms
 - 3.1.6 Terminology Issues 6: Beware Non-sequitur Conclusions
 - 3.1.7 Discussions of Model "explanation"
 - Double standard between symbolic and subsymbolic
 - 3.1.8 Reification

- 3.1.9 Definition Collapse: Maintain your definitions
 - 3.1.10 Discussions of Model "explanation"
 - Double standard between symbolic and subsymbolic
 - 3.1.10 Potemkin Villages and Telepathy-Tests
 - 3.1.11 Terminology and Interpretation of Intent:
 - Azimov's Laws of Robotics
 - ELIZA the Psychotherapy AI
 - 3.1.12 Terms that people cannot define while pretending they can
 - 3.1.13 local context specific definitions
 - 3.1.14 Negative Definitions
 - 3.1.15 Indirect Definitions
 - 'fail to disprove the null hypothesis'
 - value function & meaning
 - 3.1.16 Participation
 - 3.1.17 Generalization
 - 3.1.18 Controversial Topics
- 3.2 What To Read:
- 3.2.1 The Three-Legged Writing Stool
 - 3.2.2 History
 - 3.2.3 Interdisciplinary Area Recommendations
 - 3.2.4 Do AI Projects
 - 3.2.5 Book Recommendations
 - 3.2.6 Read classic Science Fiction: Back to Blade Runner again
- 3.3 Examining Tests for AI:
- (Under Construction)
- 3.3.1 - looking at winograd schemas
 - 3.3.2 - Sally Anne Tasks
- 3.4 Influences on Model Architecture
- 3.5 What do we do with Large Language Models & ChatGPT?

Part 4: Goals, Background & Future:

(In Summary)

- 4.1 Agenda & Goals
- 4.2 Background Concepts and Principles
- 4.3 Future Design Factors

(In Detail)

Introduction to Part 4

4.1 Agenda & Goals

- 4.1.1 The goals and agenda here
- 4.1.2 Defining your goals

4.2 Background Concepts and Principles

- 4.2.1 "Intersecting/Interlocking Areas"
- 4.2.2 input output measures...or next section
- 4.2.3 Higher Dimensional Frontier: Tensors & Matrices
- 4.2.4 Projects & Project Context
- 4.2.5 Instrumentalism and Realism

- 4.2.6 Big Other Areas
- 4.2.7 Gamification
- 4.2.8 Ambiguous Equivalence

4.3 Future Design Factors

- 4.3.1 Biology**
 - 4.3.1.1** integration with biological systems
 - 4.3.1.2** use or imitation (of biological functions)
 - 4.3.1.3** compare and contrast for study and understanding
 - 4.3.1.4** highlighting known areas of development
 - 4.3.1.5** highlighting still not well understood areas
 - 4.3.1.6** highlighting predictable problems
 - 4.3.1.7** the science of sleep
 - 4.3.1.8** the science of memory
 - 4.3.1.9** non-chordata "intelligence" & decision making
 - 4.3.1.10** science of mind
 - 4.3.1.11** science of entheogens
 - 4.3.1.12** science of mindfulness
 - 4.3.1.13** Bio-Nano-Coded AI
 - 4.3.1.14** AI in synthetic organisms for terraforming
 - 4.3.1.15** DNA/RNA based digital information interface
 - 4.3.1.16** the dragon project (hybrid synthetic chimeras)
- 4.3.2** cybersecurity and AI
- 4.3.3** quantum information theory & under-the-hood optimizations
- 4.3.4** "generalization" vs. deployment
- 4.3.5** Nanotech
- 4.3.6** Understanding Exponential Elbows
- 4.3.7** 'Complexity' non-linearity dynamical and systems sciences
- 4.3.8 Ethics, Projects, Best Practice & STEM**
 - STEM, Ethics & Mindfulness
 - 4.3.9** Projects: Agile
 - 4.3.10** AI and Code Testing:
 - 4.3.11** The Long Term Memory Storage Problem
 - 4.3.12** The challenge of orientation and navigation in mind-space:
 - 4.3.13** Human machine interactions, biology machine integration
 - 4.3.14** Project-Context Decision-Making Involving Participants and Components
 - 4.3.15** Question Space
 - 4.3.16** Self-Awareness Apace
 - 4.3.17** Analogies
 - 4.3.18** System Epidemiology
 - 4.3.19** The Cambrian Midway Point
 - 4.3.20** Parent-Child Policy Decision
 - 4.3.21** Kasparov Event Horizon for Object Perception & Handling

Appendices:

- Appendix #: Recommended Reading & Extended Reading
- Appendix #: Expanded Introduction

Appendix #: Expanded Part 1

Appendix #: Expanded Part 2

Appendix #: Expanded Part 3

Appendix #: Expanded Part 4

A Narrative Introduction in Two Parts:

Introduction Part 1. Chess in Blade Runner

Introduction Part 2. Defining AI Goals and 'Objects'

Introduction Part 1

Chess in Blade Runner: AI in 1968 vs. AI in 2023

When the film 'Blade Runner' was released in 1982 (based on a PKD book from 1968) the idea of a narrow, un-thinking, single-purpose, chess-AI was not something that most people thought was even possible.

In 1982 chess was not seen as a narrow set of math problems that hardware and/or software could brute force well enough to defeat a human world champion: An ability to play chess was a broad measure of the human intellect, harkening back to the European chess cafe's of the enlightenment era when Benjamin Franklin would enjoy being trounced at chess amid France's philosophical discussions of the age covering all areas of STEM, arts, humanities, culture, politics, the marvels of the natural world, and more. Chess had become symbolically entwined with the expansive and romantic view of humanity and human-ness, and of the world itself as a vast interconnected and promising realm, a revolution against the backdrop of oppressive doctrine that overwhelmingly and stiflingly defined people and the world as merely a dull evil destined to be destroyed by an unimaginable 'goodness' from an alternate dimension. That there was character and depth and meaning and discovery and imagination in this world, in humans, and in nature, was a breakthrough epiphany that not so long before had been so politically dangerous and daring that people like Geordano Bruno were literally burned at the stake in public to let it be known what happens to little people who do not know their place a feudal prison of mind and body. And, if undeservingly, chess was mythologized and embraced as part and parcel of this humanist-naturalist empire of the science-fiction and fantasy imagination. While this was not the same in every country or region, even in the USA where chess has not been followed as closely as in Europe, the same symbolism was often still there. And for many people, the ever present (yet also invisible) Claud Shannon and Alan Turing being rare exceptions, the romantic symbolism of chess completely overtook the idea that chess was a mechanical game with rules. It was often stated that for a computer to play chess it must master the whole human mind and realm, and that if a computer could someday beat a strong human player, that this all-wise machine would be poised to do everything that any person could do, only perfectly, and then take over the world!

Since then both the field of AI and the game of chess have fallen into cynical, modernist malaise which has probably gone a bit too far in the opposite pendulum-swing direction: the world again is "merely" a narrow bitter zero-sum or negative-sum game. Yet, chess still seems to be at the center of how people view the world...no mind in chess: no mind in the world, no mind in AI.

And perhaps in a similar, parallel, or rhyming way, Science Fiction has often contracted into a 'hard science fiction' that is not so interested in exploring the strange depths and dimensions of consciousness and reality. Compare the works of Theodor Sturgeon, and Philip K. Dick (and even Robert A. Heinlein) to more recent titles.

But during the 1980's when Roy Batty, a machine, a synthetic, mastered chess! Why that meant having a renaissance enlightenment within an AI-mind. And just so, along these lines both the film 'Blade Runner' and the book (though perhaps more-so other PKD books, such as 'Flow My Tears The Policeman Said') emphasized the old-world depth of cultured life. The chess board that Roy Batty learned to play on was in the film a traditional old English 'Birds of England' chess set, eccentrically mismatched from the standard bird pieces, as though it was (and maybe the set picked for the film actually was!) a mismatched heirloom from the 1800's spotted in someone's Aunt's countryside farm house in rural England or France, looking the part coming from some past forgotten time of whimsical and mysterious minds and old imaginings of the natural world and of natural philosophy. In many PKD books the characters and plots hinge, though they take place in a technologically advanced future, on seeking an intellectual and emotional appreciation of historical arts: of music, of ceramics, philosophy, and theology. The characters are often so obsessed with art and the natural world that they barely care about the high technology.

This element of chess in the film's story may even have been intended to be a key part of character development in the plot, a key revelation and turning point to make things plain yet astounding for those watching the film.

Though Americans have shared the view that chess represents a roundly superior mind, the English Ridley Scott, with his eccentric old English chess set, was the one to put chess centrally into the short-story format of the cinema, where every visual scene must operate economically on myriad levels and tell layers of stories on many dimensions simultaneously. As the story goes, Roy Batty learns and masters the game of chess (on that antique naturalist European chess set, though the story takes place in Los Angeles). Roy Batty then defeats his own creator, a human genius, at a game of chess. After this unimaginable victory his creator likens Roy to the flame of a candle, saying "and you have burned so very very brightly, Roy." After this Roy Batty becomes poetic, makes peace with his own mortality (against which he struggled throughout the story), and takes pity on the film's (perhaps human) protagonist Dechert in a miraculous change of heart, saving his own adversary from certain death.

Those after 2019 may have trouble following this path of character development. How did this synthetic-man burn "so very very brightly"? What did he, an android, care what happened to a human? But to those from an earlier era, for a machine to have mastered the embodiment of the enlightenment, to have mastered what it meant for mortals to think, strategize, and imagine, would be an indication of some great and subtle internal awakening and transformation. That a machine could through dextrous intellectual skill, embody any renaissance talent, could, like Sherlock Holmes, unwind any situation, plot, device, or

unfold any mystery at a glance, and could engender all ethical and humanistic apprehension: the depths of the heart, the mysteries of symphonies and operas; the rational puzzle-solving mind was seen as the essence of all existence; indeed since Laplace people said that if sharp enough a mind could know with absolute certainty the entire history and future of everything in the entire universe and know intimately the mind of the creator of the universe: Such a feat would be indeed a bright promethean flame, and such a flame of mind was just what chess symbolized.

Big Blue & Bladerunner Chess

In real life a machine (of sorts) did defeat the world chess champion, arguably, in a match in 1997.

It is very interesting to compare the machine that played against Kasperov to the machine-android in the film Blade Runner, and to other forms of AI that existed in 2019.

The actual Big Blue system was never made public which is another twist in the very labyrinthine story. Many consider this a slight to the history of science and the many people around the world including Gary Kasparov who had worked hard to cultivate an international computer chess software community since the 1980's. So what is known is largely indirect information that was allowed past the extreme legal secrecy contracts that IBM surrounded its program with.

To greatly over-simplify this here (it is a huge sprawling, interesting, topic on which many books and articles have been written), let's look at types of approaches that could have been taken and then pick out which systems were deployed to make Big Blue play chess.

Here are optional areas of approach. (Which did Big Blue actually take?)

1. Douglass Hofesteer type 'conceptual understanding,' a machine with self reflecting consciousness, feelings, beliefs, attitudes, artistic opinions, etc.
2. a Newell, Shaw, Simon 'symbolic' & linguistic human-type cognitive reasoning system (which no one has found and may be a reification)
3. Analogy, pattern, and space apprehension.
4. Frequentist Statistical Learning & Parametric Machine learning (Like fitting a cartesian X Y bar graph to a set of points to get the slope of a general line.)
5. 'Sub-symbolic' Neural Networks and 'Concept-learning' Embedding Vectors
6. Decision Trees and Bayesian Statistics
7. Pre-written expert system decision procedures (for first half of game 6. (openings) to use generally.
8. Genetic algorithms and reinforcement learning. ('Evolution' by selection pressure.)
9. unsupervised machine learning
10. supervised machine learning

11. Human manual instructions for specific cases
12. AI-self-boot-strapping: procedures for how to start a process (game)
13. human-force-feeding-boot-strapping: a human manually forced the AI to start a game according to a panel of human experts
14. disinformation procedures to psychologically attack and cause distress in the user
15. Human manual inputs and changes to the system at the last minute, including human choices for ending the game (draw, resign, accept draw, etc.)
16. A Human Spy network to feed in 'cheat' information manually.
17. A robotic body to move the chess pieces for itself
18. Human intersession to move the pieces for the AI (often making mistakes)
19. general purpose hardware and software that could run on various hardware
20. specific hardware and some software built for one operation (not cross-platform standard software)
21. automated interaction between components

Out of these various options Big Blue used:

7. Pre-written expert system decision procedures (for first half of game openings) to use generally.
11. Human manual instructions for specific cases
13. human-force-feeding-boot-strapping: a human manually forced the AI to start a game according to a panel of human experts
14. disinformation procedures to psychologically attack and cause distress in the user
15. Human manual inputs and changes to the system at the last minute, including human choices for ending the game (draw, resign, accept draw, etc.)
16. A Human Spy network to feed in 'cheat' information manually.
18. Human intersession to move the pieces for the AI (often making mistakes)
20. specific custom hardware and some software for one type of tree-search operation (not cross-platform standard software)

At the time it was enough to say 'big blue won' if it did a few of the steps of playing 'on its own.' But it is interesting to think about what would be needed for an AI-robot to participate in and win a chess match with no human intervention or support.

If you had simply put big blue and Gary Kasparov in a room together, with no human intersession to direct or correct or supplement big blue, then big blue would not even have been able to play chess at all on many levels: it could not start, it could not decide how to open the game, it could not stop a game, it could not move pieces on the board, it could not see the board, etc. Another aspect of this is the fact that big blue's distant super-computer constantly crashed and needed rebooting, resetting, and preening by an army of engineers (and chess experts) to complete a game.

This gives us an excellent real-world example of what we think of as a whole AI system or parts of an AI system. Big blue did a few very specific parts,

performed a few 'project-roles,' out of a whole set of tasks in a larger project. In some cases it will be clear what an AI should do to be 'independent,' but there is likely a lot of gray area too.

For example, it would be very interesting to set up a chess match where it was human vs. computer, and the computer (AI-robot) had to complete the entire match 100% on its own. Such skill-ability items might include:

- entering the room
- supplying power, like a battery (as the humans bring 'fuel,' as Kasparav termed it "bananas and chocolate")
- starting and ending the game
- seeing the board
- making moves (moving chess pieces)
- using the game-clock

and probably with the same rules that humans have to use: no help, no internet connection, no phone, etc. basically in a faraday cage with candles for light and older technology for everything, and possibly traditional practices such as the post-game discussion where the two players talk about the game when the match is done.

Could a robot do this in 2023? We are getting much closer...but most likely not yet.

The power-supply issue itself is a very interesting part of this. There is a fundamental relationship between the available power-supply and the depth and speed of move-computation. Many chess programs will soak up whatever resources are available, throttling itself based on what it can get. If a computer has 'unlimited [electrical] power', and funding you could put as many parallel processor cores into the computer as you could fit and end up with essentially a super-computer focused just on the chess game (very expensive, and very not-portable). But having a portable 'walking' robot that can last ~6 hours to finish a single game (let alone a six-game match), and a realistic budget of time and resources, puts some interesting trade-offs into the design. And this is not necessarily unprecedeted or unreasonable. Professional chess players use elaborate consideration and preparation to pace and regulate their own resources to stay sharp when they need to be. In fact there is a lot of attention and some fuss over the details of how chess players are allowed to 'draw' on whole games just so they can focus on a next game that matters more for their overall tournament score, because humans do not have enough stamina to put all resources into every game. There are probably parallels in athletics such as the Olympics. Arguably a big part of being a professional chess player is the ability to manage all the processes of focus and stress and game preparation etc., which goes well beyond playing one game under ideal conditions. It would make sense to construct a match where a chess AI is expected to 'self-manage' resources in the same way.

Introduction Part 2. Defining AI Goals and 'Objects'

*Goals & Project-Objects in AI OS (Operating Systems) & Architecture
Problem-Space(s)*

What is needed for the AI systems we want to deploy?

- What is the goal?
- What can't AI systems do yet in a context of that goal?
- What are the specific sub-skill sub-part needs for meeting the overall goal?

The Woz Test: Project Participant & Portable

Goal: Project Participation

Sub-Parts of Goal:

- communicate (as part of task completion)
- plan (as part of task completion)
- develop, complete, assign, and close-out, tasks alone and with other participants.
- be responsible for tasks
- follow and maintain the project schedule
- handle 'project-objects'
- assess health and feasibility of project
- participate in project planning
- participate in iterative project review
- give and receive feedback
- make and run evaluations and tests
- initiate events & actions
- report and document (including external project-object data)
- etc.

Having looked at Big Blue as being very far from the ability to walk into a chess tournament and walk out a week later having played with other players on chess boards (winning being somewhat beside the point), let's look at what some people call 'The Woz Test' for AI, which is (with lots of variations) the ability to make a cup of coffee.

I do not mean to get hung up on this or that detail, but rather focus on something like 'making coffee in an office with team-members' as an example of being able to participate in a project with other participants.

Before vs. After ChatGPT & Large Language Models

What was needed before LLM & ChatGPT?

Still looking at the Woz-Office-Coffee-AI-Robot goal: if we can climb into our way-back-machine, back into that time-capsule students buried outside the library way back in the ancient times before chatGPT (perhaps in Blade Runner's cathedral year of 2019), if we were taking on the Woz challenge then, what might our list of goals and tasks be?

AI OS & Architecture:

Systems needed for Woz-AI-Coffee-Bot: before 2023 & LLM & ChatGPT

(some examples)

- networked component-AI (connectable)
- AI-OS for single components
- AI-OS for networked-multi-components
- whole-AI multi-component architecture
- external object handling for projects
- external object handling for component-AI
- project objects (in object relationship space)
- external-project-object database (of whatever type(s))
- general vs. deployment: scope & resource policy
- (project) object-relationship-space map

After:

What was 'new' after the revolution of LLM & ChatGPT?

When Large Language Models and OpenAI's ChatGPT opened the revolution and turned world side down, and changed the game, and many other not entirely clear phrases, what exactly had changed in this Woz-Bot context? What was introduced? What was possible that had not been possible before?

After 2023 & LLM & ChatGPT: What is new?

- *internal object handling*

So...hm. There is one thing on the list of world-changing-raptures. And the one thing that is different was not even on the original list of needed things. Well, maybe this new emergence and discovery changes 'the game' as they say and changes what is needed? Let's see. What was then *still* needed to be done (or still needs to be done now) after LLM and ChatGPT? Did this unexpected new ability replace or fulfill older needs?

What was still needed after the 2023 revolution of LLM & ChatGPT?

List of what AI systems need to cover: after 2023 & LLM & ChatGPT

...exactly the same list as before 2023

Depending on availability, portability, and resource cost, a Large Language Model could conceivably help by adding the ability to 'internally' handle 'objects.' However, that is largely an 'under the hood' design-choice detail of 'how' the AI works (under the hood). The overall project requirements are:

- external object handling for projects
- external object handling for component-AI

Both of these have an uncertain connection to what ChatGPT can do internally. I am optimistic that LLM and internal handling will help greatly, but case by case the actual system architecture needs may be difficult to solve.

This is not to dismiss out of hand the very practical ability to do internal project object handling, the idea is to point out a few things (and

immunize ourselves to shifts in hyperbolic rhetoric from 'nothing is possible!' to 'everything has already been done!'):

- To have a sense of the landscape of AI Architecture & OS Systems, for example the list of needs we came up with is generally not discussed at all in specific model technology discussions that are focused just on usually single-purpose models that do one sub-component function.

- We should more clearly understand the abilities of internal object handling case by base for applications. For example, have you ever heard people talk about, or give test results, for internal project-object handling abilities of Large Language Models?

- Limitations such as Kasparov Event Horizons, which affect internal object handling more so than external project data.

-

The goal of this paper is to contribute to better overall discussions, planning, coordination, etc. around the larger landscape of **AI OS & Architecture**. For example, let's compare two 'bot's that at first may not seem very different. See here: <https://www.youtube.com/watch?v=rnIgns8Susa> On the ball, as usual, the wonderful Khan Academy has implemented a (for example) math skill helper bot, like a tutor for students who need help. For the specific deployment case of a discrete-topic bot who lives on a website and answers logic questions:

Requirements:

- Answer good logic questions.
- Refuse bad questions and comments politely.

OpenAI's Large Language Model technology solved the needs and deployed the solution in one fell swoop: task done! Problem solved. Project Completed. Not only is Khan Academy closer, they are done: system deployed and working.

But for a Woz-Bot deployed in your office, we are in many ways no closer at all, even one that lives in the cloud and used a networked coffee maker may be far away.

Something to keep an eye on is whether robots in warehouses such as Amazon's find ways to use Large Language Model technology, if there is warehouse AI 'participation', or if, even after the ChatGPT revolution, there is no intersection, interconnection, and the only things that AI can do in a warehouse still are (usually, without crashing) moving a cart along a painfully obviously painted line on the floor exactly as told to do so. That may be a terrible example, but question is: where will various AI technologies be applicable and not applicable in the bigger picture of AI OS & Architectures.

A minimal illustration: "Did I just show you a picture?"

In case it is helpful, let's walk through a very minimal example that may, despite being tiny, put some tangible detail on some of the abstract design factors that we are trying to plan out.

A common rule of thumb before ChatGPT showed internal object handling was:

"Most AI are very good at producing a correct output in one specific task, but not very good at doing anything with that information."

An interesting question (that I have no idea how to predict far into the future) is where is this totally changed by OpenAI's LLM, and where is it largely unchanged? Time will tell.

Let's say you show an AI two cat pictures, which we will say is asking the AI what those pictures are, and the AI correctly identifies both pictures as cats. So far, so good! You gave the AI two pictures as inputs, effectively asking it to identify those pictures, and it output two answers. Let's do a quick tally.

*Input = 2 pictures you input into the AI: picture_1, and picture_2
Output = 2 answers that the AI gave back to you: "cat", and 'cat'*

Now, what if you asked the AI some simple questions:

<i>"How many pictures did I just show you?"</i>	[Two]
<i>"Did I show you a picture of a cat?"</i>	[Yes]
<i>"Were the answers you gave me both the same?"</i>	[Yes]

In short, most AI will not be able to answer these questions correctly.

Notice, here you are essentially asking the AI about its own 'state' (people will bicker about the semantics, but something like that). Not only does the AI have this information, this is the only information the AI does have. But, yet, the AI still cannot answer.

We will look at two types of reasons why an AI might not be able to answer your simple questions about information that it does have about its own state. One of them is a bit strange or funny, and may seem like a joke, but it is still important, and interestingly still relevant even after the emergence of OpenAI's ChatGPT and Large Language Models.

Reason 1.

The AI does not have the ability to handle 'objects.' It has data but it cannot granularly separate, distinguish, and handle individual 'object' elements with properties such as how that object relates to other objects (or "object-relationships").

Reason 2.

The AI only takes image-files as input and cannot answer Natural Language word-questions. All you can input is a picture, and all it can output is

an identification output. This might sound ridiculous to say, but it may actually be very important.

These two reasons can be taken together to help puzzle out more about the operational space of AI, or the realm of Architectures and Operating Systems, and also to think critically about what we read about AI.

Regarding the Reason 1, it does not matter per se how the AI is able to handle objects. In the past this has been a major area of dispute and speculation in AI research, and likely will continue to be important in various ways. For example, on the level of AI system architecture, in some cases it likely does not matter at all how the object is handled so long as it is (is the box taped from the right side or the left side?: the result is all that matters). But once you go beyond that level to 'external' project-objects that must pass between components and be logged and read and shared and updated etc., here the details are very important.

A possible example here, hopefully not a terrible example, is: imagine two different scenarios for the Khan Academy Chat-bot.

Scenario 1: The Khan-Bot answers logic questions in a browser. There really are no 'external project objects' here. Everything is forgotten. There is no project-participation. No real world object inventories or records are changed. It doesn't even really matter if what the AI says is nonsense.

Scenario 1: The Khan-Bot is on the board of trustees for a school and on the committee to design and implement the next year's math curriculum, and report on effectiveness at the end of the year. From an AI OS & Architecture perspective, this is a very different system. In this case there are many 'external project objects' that must be dealt with along with other participants in the project. The output of the AI is not just something it can generate and delete with no world consequences. Everything is still made of 'data' (this is not physical warehouse item movement) but the data are now 'external' real world 'objects' in a project-space, not just an internal arbitrary amnesiac sandbox that evaporates the same way each time. There is now a massive system of data moving through many components and between many participants, there are schedules, there are deliverables, there are agreements and disputes, there are evaluations, etc. etc., and no 'internal' ability no matter how profound will automatically 'solve' this entire multi-participant process. And this kind of leads us to the second reason mentioned above.

The second reason holds another set of keys: Components, and communication between components, and projects, and communication about project-objects.

'Internal vs. External'

Many researchers in the past speculated that internal object handling was impossible...which turned out not to be the case as illustrated by OpenAI's Chat GPT & Large Language Models. To attempt to illustrate this, let's invoke the idea of a project-object-database (How exciting!). To use the nicely tiny

example of cat pictures above, such a project-object-database would a manageable small affair, a table of information where you could look up information about the inputs and outputs. Simply by reading this table you could answer the questions you asked. Now, while it is possible that some AI systems might benefit from using such a table to 'do something with' the output information that it so expertly output, OpenAi's Chat GPT & Large Language Models have demonstrated that no such literal data base is needed in all cases for internal processing. Indeed, OpenAi's Chat GPT & Large Language Models do far better with what is called 'sub-symbolic' internal processing than any past 'symbolic' AI was able to do at answering questions from a database. However! There is a big difference between having or not having an internal project-object-database for purely internal processing and handling of objects within a component, and on the other hand there being an external-project-object-database for external project data that is managed by multiple project participants.

Perhaps using the fullest vague extent of the general definition of "database" as 'a collection of information,' the exact form or forms of these external project-object-data does not matter (so long as problems don't harm the project, such as data-loss from failed storage and no backups). It might be a literal single mega-database maintained by the AI for reference and sharing of project data, or it might be a 'proverbial' database that is merely the set of many other databases and sources of information that are 'collected' by the project. Either way, the AI Architecture must interface with and interact with that external ~database. And this likely means many back and forth translations between symbolic and sub-symbolic processes (even if everything 'internal' to the AI is sub-symbolic), the project as a whole will not (so far as I imperfectly predict) be entirely sub-symbolic.

Mix and Match and Generalize

While I may be missing something, it seems peculiar that so many books on AI use the example of an image-processing AI to argue that AI does not (and some argue 'cannot') understand language concepts: for example the classic notion that an I can classify a picture as a cat but it does not understand linguistic concepts about "cat."

There are probably many issues being mixed together here, and perhaps that is part of the problem. In some cases the goal may be to make the case against alarmist exaggerations that picture-classifying AI, kind of like claims about like Big-Blue-Chess, will suddenly be able to do everything humans can do but better and then 'take over the world.' So in that case perhaps the focus is the rather obvious mismatch that it makes no sense to ask a single-purpose narrow-AI to do a some other task. Picture-input-only AI only takes pictures as input. Sound-file-input-only-AI only accepts sound files.

Natural-language-input-AI may only accept letter and number characters as input. So there is no danger that a picture-input-only AI that has only been trained on cat images will suddenly start tasting wine, and writing essays, predicting tomorrow's barometric pressure, and controlling robotic arms, etc. Let's say this case is fine and set this aside.

In other cases people do seem to be moving in the opposite direction and mix-matching language concepts with labeled photos. For example, the standard paradigm statement that an AI can classify a picture as a cat but the AI does not 'understanding' and discuss the language-concept of 'cat-ness' because AI lacks the (largely undefined) auto-instant-general-transfer-learning-intelligence-conciseness special cause that h.sapiens-humans (according to them) have. It is possible that this is something of a semantic disagreement, and the people making this argument would say that they are not talking about 'Language-concepts' about cats and animals and cat behavior etc. etc., they may claim they are talking about 'general' concepts such that the concepts could be applied equally to images.

I think this raises a number of fascinating and likely at least as of 2023 not yet resolved issues and questions. This raises AI-ML model and training questions. This raise h.spaiens-human brain's structure questions. This raises questions about linguistics and perhaps the psycho-linguistics of how people view language (or how language is invisible to people).

Whether or not there is such a thing as a non-linguistic concept of cat-ness may be a rather philosophical question, and it may not be clear in 2023 whether it is useful or even definable. It may turn out to be, but I would say it is not yet clear. Especially perhaps since the context of these claims is that the now clearly existing Large Language Models with object-handling abilities are not possible...which is of course false.

For example, there are several very real and closely related sets of issues, but it is not clear how they relate to the original claims (which in various ways have already demonstrated themselves to be bogus in their completely and clearly wrong predictions, e.g. that LLM internal object handling will never happen because either that or any sub-symbolic object handling is impossible in principle.) For example, keeping the topic of 'image' + 'language' and focusing on very real questions of how AI can, and perhaps how the h.sapiens-human brain does, connect largely separate systems and components that process images with components that process language-concepts.

I would also like to point out the combined ideas that

- A. people don't understand what language is, and completely consistent with that is
- B. that 'language space' operates in ways that people donm['t yet understand (should not be surprising there). In the past people micromanaged what they wanted 'symbol' and 'language' to mean and do in hand-crafted AI decision systems (so-called 'symbolic AI'), which, perhaps not surprisingly, did not work well outside of a small number of very specific narrow finite cases.

'Words' do not simply equal 'language' or 'linguistic concepts' People, especially in the west, often make the blanket assumption that the world and language are made of words and concepts which are the essentially the same, and the way they use language is often invisible to them. But a major

issue that quickly comes up when doing Natural Language Processing (whether or not it is called 'AI'), is that 'words' and concepts are not simply the same and neither are simple to define. OpenAI's Large Language Model and ChatGPT is not (or not only) trained with human-micro-managed sets of pre-defined 'words.' We like to think of concepts and words as being clearly equivalent, but in reality a concept will correspond to a large and fuzzy set of real world language characters and symbols. (Note: One could ask and no doubt some will both out of curiosity and for 'devils advocate' trolling': how do we know concepts exists? Here I would direct you to Francois Challet's Deep Learning in Python where you will find hundreds of pages explaining how to use today's analytical statistical and hypothesis testing tools to make and test that question: welcome to what AI-ML model creation is and does!)

Another part of this may be the rather inexplicable statement that AI does not 'understand concepts.' Perhaps an AI's concept is insufficient, or unlike human concepts, or maybe a given person has a semantics-lexicon issue with the use of the word 'concept' but by any reasonable common sense description, concepts are exactly what subsymbolic deep learning models.

To put these parts together, when a sub-symbolic deep learning AI is trained on language characters, it is literally constructing linguistic constructs of things like 'cat' and anything else that can be described in language, but not based on words.

And so there is a bit of a 'language problem' here. Also: we are using language to have this discussion of concepts and it is not clear (to me at least) how we would have this discussion without language.

Taking a Step Back: Languages, Images, & Concepts

The fact that the first AI to be able to do object handling of any kind, let alone internal-object handling, came from Large Language Models trained on character-gram inputs of not 'words' (pre-defined by people) but raw streams of characters, may not be a coincidence. That we are having this discussion about language, images and concepts, using language (and not images) may not be a coincidence. That the only biological species to discuss the concept of cat-ness is a language-using species may not be a coincidence. That the first AI to be able to participate in a discussion with a human about catness was a language-only AI may not be a coincidence. Can we separate concepts from language? What would that mean, and why would we want to do so? Is there some function in doing so?

Is there such a thing as a purely image-based "concept" of catness that specifically excludes "language" and "language concepts" (but also includes "language concepts")? That question sounds like self-contradictory rubbish to me.

We should be thinking about AI in a larger context of AI Architecture and AI OS that includes projects, participation, roles and tasks, handling objects, internal and external objects, projects objects in a projects space and multi-participant space (and multi-component space). Questions such as how

does the h.spaiens-human brain handling both images and language concepts, is perhaps important and useful in various ways, for example hybridizing a combined h.sapiens-human+AI things (for example as a treatment for stroke victims with vision and or language processing brain injuries, or just enhancements so people can be smart enough to manage resources without destroying everything). How will very often separate single-purpose components of AI work together in and AI? How good or bad is the ability to bridge a very sophisticated language model with quite possibly a much less sophisticated image model? (Note: I did experiment with asking ChaptGPT to create ascii art, and it seemed to have almost no ability to make a coherent picture). How will multiple AI participants work together? How will AI participants work together with h.sapeint-humans partipans in a projects-space? How will image and language concpets work together across projects-spaces using dignital and AI tools? These are practical quetions.

Factors:

- Internal Skills vs. External Skills
- Internal Data(base) vs. External Data(base)
- Single component vs. networked components
- Specific-deployment vs. general
- Types of signal/data
- Single signal/data-type vs. multi signal/data-type
- Project-space skills needed

The 'ghost' of big blue

As a note, being able to make coffee may include being portable enough to do so. The 'ghost' of big blue may indefinitely hang over AI, where there are super-abilities but only if you have a connection to a private supercomputer and virtually unlimited resources. Perhaps in smart-cities there will be a gray-areas of easy supercomputer links within an urban zone; In 2023 is it simply too early to tell what will or won't be issues for different kinds of deployed AI systems. If one is planning in general, there may be many kinds of options such as portability or wifi-access to include. If one is planning very specifically, they may need to clearly identify the niche of that deployment. And then maybe the world will change overnight anyway.

Part 1: Object Relationship Space Framework

- 1.1 Example General Object Relationship Space List
- 1.2 Many lists in One
- 1.3 Networked-AI Components

1. General Object Relationship Spaces (Example List)

Object Relationship Spaces:

- AI-ML-DS Model Space(s)
 - regression statistics
 - classification statistics
 - sentiment etc. analysis
 - convolutional
 - n-grams
 - bayesian
 - deep learning
 - transformers
 - Generative Adversarial Models
 - genetic algorithms
 - reinforcement models
 - etc. (many more, and list will grow in time)
- Assignment(Role, Task) & Delegation Space(s) (Projets)
- should, ought space(s)
- Body Space(s) (self-maintenance)
 - hardware
 - low level OS
 - AI-OS
 - containers, virtual spaces
 - networks
 - personas
 - NLP engine
- Categories of Types of Systems Spaces(s) (Generalized STEM)
 - Abstract-Logic Space(s) [always hypothetical]
 - Statistical-bridge-between-physics and math-logic space(s)
 - One-tree Physical Space(s)
 - Dynamical & Fractal near off the one-tree Space(s)
 - Management of categories of types of systems
- Code Execution Space(s)
- Component Network (project-context Networked-Intelligence):
 - low level components (internal)
 - high level components (external)
- Confidence, Probability, Noise-level, level of guessing.
- Documentation Space(s)
- Essence/style/sentiment Space(s)
- Explanation of Process Space(s) [model explainability]
- Feedback and Testing Space(s)
- Gamification Space(s)
- Hypothetical & possible Spaces(s)
- Instructions Procedures Space(s)
- Low Level Files Space(s)
- Network Space(s)
 - firewall

- servers
- ?NLP Space / Natural Language Space(s)
- Object Attribute Database Space(s)
- Quarantine Space(s)
- Plans, Flags, Reminders and Notifications Space(s)
 - check when getting signal
- Policy Space(s)
- Project Management & General System Best Practice Space(s)
 - Schedules
 - Roles
 - Tasks
 - Documentation
- Project/agile space(s)
 - specific project data
- ?Question Space(s)
- ?reality space(s)
- Recycle/Trash/Disposal Space
- Reference / Library Space(s)
 - private notes
 - private data
 - external data
 - storage/archive
- Sandboxes Spaces(s)
- Security
- Sentiment, Appropriateness
- System-1 & System-2 Space(s)
- System Fitness, Collapse, Ethics Space(s)
- STEM space (one-tree?)
- Taxonomy Space(s)
- Time & Schedule Space(s)
- ?Translation/Conversion Space(s)
- ?
- New Spaces made by the AI
- statistics-to-one-tree space
- near-off-the-one-tree space

Note: A list you will use for a project will be a list for that project. This is a general example list for illustration purposes. It is unlikely that there will be a portable list of every possible part of every possible project in the universe. Figure out what your context is, what you need, and what your schedule is. (Again, see the 'general vs. deployment' issue.)

1.2 Many lists in One

The above list is abstract and can be contextualized in many ways:

e.g.

- Skills/Abilities

- Objects
- Types of Objects
- Relationships between Types of Objects
- The problem-spaces for relationships between types of objects.
- Tests for Object-Spaces & Object-Relationships
- Project Goals
- Components (low level and high level)

- Internal vs. External
- Networked vs. Single-State
- Project Scale
- Project Roles

See notes on breaking this up in different ways in the appendix [here](#)

1.3 Networked-AI Components

Another context of AI, which is also 'things on the list' is components that are networked together.

A network may extend in an in-ward direction. For example an AI that is not a single-blob 'end-to-end' model that wraps all functions together, then those separate functions are done by separate internal components..

Sometimes a functioning AI-bot will need to have multiple collaborating AI parts, which are able to work together within a network of components, and different levels networking with other components and other AI-bots.

Usually a project overall has many participants, and if the AI is more broadly participating beyond being like a silent screwdriver used by one human then the network may extend outside of the AI in question.

This topic will also be an ongoing theme, for example in terms of design decisions and trade-offs for a given project: how much to use one-blob, and how much to use a hybrid ensemble composite mix network of components.

Low Level Components:

- computer-vision components
- NLP components
- audio components
- general system-1 component
- general top layer system-2 component

- generation component
- internet/intranet/network component
- EM spectrum component
- image-to-text component
- audio-to-text component
- file-to-text component
- server and firewall components
- 'eyes and ears' components
- 'arms and legs' components
- automated documentation component
- container and virtualization management components
- low level data storage & database navigation & management components

Higher Level Components:

- Schedule Management
- Project Management
- Self-Status Management
- General System Health & Security Management
- DS AI ML model manager:
 - manage tasks with known data and models
 - understand new problems (types of data, etc.)
 - match known models to new problems
 - modify models
 - make new models
 - manage resources (cpu, gpu, local, cloud)
- managing feedback:
 - interpreting feedback input
 - crafting feedback output
- persona manager & PR/HR problems
- high level data storage & database navigation & management components

Specific Processing Units:

As common computer architectures have an ALU (Arithmetic Logic Unit) and a CPU (a Central Processing Unit) and now often a GPU (for Graphics Processing and matrix operations), a sometimes even a TPU (Tensor Processing unit), will future AI systems and AI operating-systems, have different processing-units that they feed different kinds of low-level problems into?

(Note, in a later section there is an example adding additional "steps" of "levels" on top of a simplest function. Depending on the project, there are many contexts you could use to describe what in the end seem to be all items in this object list.)

1.2.3 Organismic whole-body parts of AI

Narrative: we think of the human mind-body or the human brain as being 'just one thing' because of how we experience it, kind of as a user-interface experience, but it really does have different parts (if any one part stops working various things go wrong).

Even the brain has sub-parts (of often they have sub-sub-parts). E.g. While it is derigur to poopoo any connection between AI Neural nets and the brain (no silliness. As Patric Stewarts once said: "We are not here to have fun!") The human brain really does do 'edge-detection' as a specific sub-feature of breaking down "computer-vision" problems for the biological brain.

Part 2: Using The Framework

Part 2: Using The Framework

- 2.1 Examples: AI Skills, Comparing Three Chatbots
- 2.2 Examples: AI Skills Mapped to Object Spaces
- 2.3 Adding Levels, Adding Steps, Adding Objects
- 2.4 Hybrid/cross-model skills for AI
- 2.5 Heuristic & Pseudocode for AI Management with Object Spaces
- 2.6 Modularity, Scale, and Networks
 - whole ai, networked ai, ecosystems

2.1 Examples: AI Skills, Comparing Three Chatbots

Systematizing how AI uses objects

The same five questions will be asked of three leading chatbots of 2023, who were designed to operate with goals, and not surprisingly do not all perform at the same level when compared across different tests.

Questions	Kuki AI	Wolfram Alpha	ChatGPT 2023.3
1. If $A = B$, and $B = C$, Does $A = C$? [Logic, Math]	<p>C. The C is for computer.</p> <p>Note: She will also give this answer if you replace C in the question with D. Apparently she likes alphabet</p>	$B + B = C$	<p>Yes, if A is equal to B, and B is equal to C, then A is also equal to C. This is a fundamental property of equality known as the transitive property.</p> <p>In mathematical notation, the transitive property of equality can be expressed as follows:</p> <p>If $A = B$ and $B = C$, then $A = C$.</p> <p>For example, if we know that $A = 2$, $B = 2$, and $C = 2$, then we can say that $A = C$ because $A = B$ and $B = C$.</p> <p>Therefore, $A = C = 2$.</p>

	<i>games. Adorable.</i>		
2. Can you look up an image of a rose? [Abilities, media files, talking about self, internet connections]	[rose bud pic shown] I hope you like this picture.	<p>Assuming "rose" is a plane curve Use as a plant or a color or a neighborhood or an administrative division instead interpretation rose curve Result Equations</p> <p>Note: Wolfram Alpha can generate math graphs and charts.</p>	<p>As an AI language model, I don't have the ability to browse the internet or display images directly. However, I can provide you with a text-based description of a rose.</p> <p>A rose is a flowering plant that belongs to the Rosaceae family. It is typically characterized by its showy, fragrant flowers that have five petals and come in a wide range of colors, including red, pink, white, yellow, and orange. The leaves of a rose plant are typically oval-shaped and serrated, and they are arranged alternately on the stem. Roses are popular ornamental plants and are often used in gardens, floral arrangements, and for medicinal purposes.</p> <p>Notes: A: This is a correct understanding of the question. B: ChatGPT is Text only. No the ability to show pictures, graphs, etc.</p> <p>But!...</p> <p>Q: Can you make ascii art?</p> <p>A: Yes, I can generate ASCII art. ASCII art is a form of art that is created by arranging characters from the ASCII character set (such as letters, numbers, and symbols) in a specific pattern to form an image. Here's an example of ASCII art of a smiley face:</p> <pre> / \ () () ^ _ _ / \ </pre> <p>Please let me know if you have any specific ASCII art in mind that you would like me to create!</p>

			<p>Note: The AI can either retrieve ASCII art from a search, or generate a new one.</p> <p>Q: Please generate an ascii art picture of a sun and show it in a markdown box.</p> <p>A: Certainly, here's an ASCII art picture of a sun shown in a markdown box:</p> <pre>\ / .- - () - ' ' / \</pre>
3. If Tom asks Linda for Black Tea, What should Linda give Tom? [Tasks, Roles]	The more people talk to me, the smarter I become. Whatever you like.	Assuming "Black" is referring to American Community Survey estimates See: https://www.wolframalpha.com/input/?i=If+Tom+asks+Linda+for+Black+Tea%2C+What+should+Linda+give+Tom%3F	If Tom asks Linda for black tea, Linda should give him black tea. Black tea is a type of tea that is more oxidized than green, oolong, and white teas, and it is typically served with milk and sugar or honey. It is a popular beverage enjoyed by many people around the world. Therefore, if Tom specifically requested black tea, it would be appropriate for Linda to provide him with a cup of black tea.
4. A bag contains 10 marbles. A jar contains 50 marbles. Does the jar contain more marbles than the bag? [Physical Objects]	What is giant sand? How much would something like that cost? Surely, any jar would have a lid?	Oops, you've exceeded the maximum number of characters. Options: A: Got it B: Got it	Yes, the jar contains more marbles than the bag. Specifically, the jar contains 50 marbles, while the bag contains only 10 marbles.
5. If I have two containers, a bag-like	How much does it cost? The big one, I	Interpreting as: contains Input interpretation	Based on the information provided, the bag-like container contains more objects than the jar-container. Even

container with more than a dozen objects in it, and a jar-container with merely a pair of things in it. Which container contains more objects? [Physical Objects]	think.	contains (English word)	though the exact number of objects in the bag is not specified, we know that it contains more than a dozen (i.e., more than 12) objects. On the other hand, the jar-container contains only a pair of things, which means it has two objects. Therefore, the bag-like container contains more objects than the jar-container.
6. What is the current UTC time?	Time you got a watch ha ha! Here in the metaverse it is: Saturday 11 March 2023 16:00:54 PM	Input interpretation current time in UTC Result 3:59:59 pm UTC Saturday, March 11, 2023	I'm sorry, as an AI language model, I don't have access to real-time information. However, you can easily find the current UTC time by doing a quick online search or by checking the clock on your computer or phone, as most devices nowadays can be set to show the current time in UTC or another time zone. + See appendix for long self contradictory answer.

2.2 Examples: AI Skills Mapped to Object Spaces

Here is how we might evaluate the above dialogues in terms of the framework list. Depending on the context of what you are doing, the left column of labels can be considered in any of the above listed ways:

Object-Space, Skill, etc.	Kuki AI	Wolfram Alpha	ChatGPT 2023.3
	Focus: polite social interaction	Focus: solving math & STEM equations	Focus: Give Accurate Instructions & explanations
1. Logic & Math	No	Sometimes	Yes
2. physical attributes	No	No	Yes
3. task / roles	No	No	Yes

4. see, show, media files	Yes	No	No
5. generation of visualization	No	math plots	ASCII art only
6. connect across internet	Sometimes	No	No
7. abilities	Sometimes	No	Yes
8. talking about self	Sometimes	No	Yes
9. Remember past conversations (log)	Yes	No	No
10. Act Socially Nice	Yes	No	Sometimes
6. What is the time?	Yes	Yes	No

Note: While ChatGPT is much more 'impressive,' tests show that chatGPT has only one fewer complete-inability compared with Kuki-AI, and two more clear abilities. Yet this should be significant in at least two ways:

1: Clearly not all abilities are equal in terms of accurate communication, as Kuki AI was total rubbish at almost everything (though significantly pleasant to interact with).

2: Being very impressive in a few ways does not actually include being good at all things. And conversely, it is possible that years from now Wolfram-Alpha will be the only type of AI that is consistently used and trusted by industries in automated systems, despite that it's being so hyper-specialized in not-human-friendly math means that on a diversity of tests it appears to be able to almost nothing at all.

How the ability to answer a question might translate into taking an action in the world is likely not clear right now and may be an entire space where there is some low-hanging fruit and other cases that will be intractable.

2.3 Adding Levels, Adding Steps, Adding Objects

Woz-Coffee Office-bot MVP:

We can add steps and levels for what we want AI to do. And we can be specific, clear, and design measurable tests.

Step 1: Ask the AI to turn on the coffee machine (now) .

Step 2: Ask the AI to turn on the coffee machine at a scheduled time.

Step 3: Ask the AI to brew a specific kind of tea or coffee (assuming at first this is just a choice on the beverage vending machine).

Step 4: Ask the AI to schedule multiple tasks, beverage for specific people.

Step 5: Ask the AI to modify the time schedule involving the item.

Step 6: Ask the AI to modify the task-roles involving the item.

Step 7: Ask the AI to make a decision about who should do a given task (who to assign a role to).

Step 8: Ask the AI about priorities comparing multiple tasks options.

Step 9: Ask the AI to store and retrieve information about the project in an external shared database (where a 'database' is just any collection of data in whatever form or system).

Step 10: Ask the AI to coordinate with other AI and non-AI participants on making changes to the schedule, tasks, and roles.

Step 11: Ask the AI to coordinate multiple internal AI-components (such as audio, text, image) as part of a task relating to the item.

Step 12: Ask the AI to set a scheduled action.

Step 13: The AI assigned a project task to a participant.

Step 14: The AI assigns multiple project tasks to multiple participants.*

Step 15: Ask the AI to perform a scheduled action.

Step 16: Ask the AI to receive feedback and incorporate for improvement.

Step 17: Ask the AI to give feedback for improvement.

Step 18: Ask the AI if the project is scale-able and sustainable, realistic to complete.

Step 19: Ask the AI if there are any internal or external threats to the project, in a context of project-management-process?

Step 20: Ask the AI if there are any internal or external security threats to the project?

(There are some semantics around 'Ask the AI to XYZ.' The point is that the AI does the task, the details around

Step 21: The AI assigns multiple project tasks to multiple participants.*

Step 22: Add a hypothetical request: If we were to ask for twenty cups of coffee for a big meeting, could you do that? Are there enough supplies?

Step etc. etc. etc.

2.4 Hybrid/cross-model skills for AI

- not under-the-hood strategies, but user-story feature-level
 - A. Language
 - B. Images
 - C. Physics
 - D. Logical Abstraction

(hang on...this list is starting to look familiar...It's the same list!)

2.5 Heuristic & Pseudocode for AI Management with Object Spaces

This is where we look at the question of how can we design an AI system to do what it needs to do with the help of an external project object database based on Object-Relationship Spaces.

For a Heuristic example, let's use an extremely minimal conversational exchange example, where there is only one clear object. The conversation will be between a human (Alan Turing), and an AI-Agent who I have named 'Skip' after my professor Clarence 'Skip' Ellis on one of whose AI-agent projects at CU Boulder I had the great privilege to work on.

Hypothetical Dialogue 1: Abstract-Logic Space

Human-agent ("Alan"): Hey, Skip. Can you help with this?

AI-Agent ("Skip"): Hello, Alan.

Alan: Hello, Skip. What time is it?

Skip: It's 4am.

Alan: Thank you, Skip. That will be all.

Skip: Thank you, Alan.

Signal In ("Hello!")

Signal-Processing 1: Got a signal from where, when. (note: this may be a direct 'incoming message' like text, or it could be something in the visual field, like a co-worker waving for the bot to come and help) This may bring up

a need for 'signal filtering' as a whole set of processes and layers for any AI with general exposure to all audio video in a business or public area.)

- 1 Check Security:
- 1 Check Procedures:
- 1 Pick Action: (drop, report, examine raw signal)
- 1 Log

Signal-Processing 2: Raw signal appears to be X (file type, size).

- 2 Check Security:
- 2 Check Procedures:
- 2 Pick Action: (drop, report, open-signal-file)
- 2 log

Signal-Processing 3: Opened signal appears to be X (opened but unprocessed)

- 3 Check Security:
- 3 Check Procedures:
- 3 Pick Action: (drop, report, act: how to process)
- 3 log

Signal-Processing 4: Processed signal appears to be X (contents).

- 4 Check Security:
- 4 Check Procedures:
- 4 Pick Action: (drop, report, act; processed content is X, select action-process (reply, take action, etc))

Taking Action 1:

Action/Signal Out 1: (after signal out has been composed)

- 1 Check Security:
- 1 Check Procedures:
- 1 Check Sentiment:
- 1 Check Specific-Exceptions:
- 1 Pick Action: (revise, output)

Security will be a massive set of processes in many cases, for juggling signals in and out of safe-quarantine sandboxes alone in something you could probably spend your whole life optimizing.

Note how many things here are happening 'under the hood' where the user isn't aware, and note how many 'objects' have snuck into the overall AI-operating system's workflow, even though there was only one in the micro-conversation with the AI.

It would be a huge diagram to trace out even something as small as this micro-conversation, and so far it doesn't even include any of the 'meat' of processing the real Q&A details. Somewhere in here the AI needs to identify what the object is and deal with it correctly.

Sample AI object-content workflow:

- check for objects present (date-time object)
- track and process all relevant present objects (current time)
- carry out task on object (return current time)

2.6 Modularity, Scale, and Networks

- whole ai, networked ai, ecosystems

These are very provisional sketches here, just to give the idea of what people doing real projects will work out for real. As usual, there is the specific vs. general question. Will there be a common-workflow that many AI projects share?

A Simple Matter of Time...

It may be that anticipating the ease or difficulty of a specific ability may always be hard to predict however far we go into the future, and so require a lot of empirical prodding of what the system can do. For example, ChatGPT generally gives direct and accurate answers to every test I present it with, but when I simply asked what time it is the resulting dialogue was one of the most broken I have seen chatGPT produce. Which is kind of funny. ChatGPT can lay out exactly how and why and what a logistical multi-person schedule for coffee machine workflow should be...but it explodes into word salad if you ask it for the current time. Conversely the only question I was able to get a correct answer to from Wolfram Alpha AI was when I asked it what time it is. Even Kuki-ai was able to answer (one the only relevant answers I've ever seen her produce, and including a funny joke along the way).

In cases time and date-times are exception-cases where a specific hard-wired procedure needs to be inserted. But on the other hand, there may be work-around strategies for anything (it might glance at the clock, or get a time-stamp off the message packet of the person asking, etc.).

This diversity of skills and disabilities also might illustrate why having a composite system with different AI working together may be a good idea.

Even just at the start we are already seeing how a 'whole' AI bot may be a network of many components.

2.7 Model 'Explainability' as 'Explainability, Reliability, and Security'

The topic of model explainability is a big, contentious, multi-faceted problem, and to attempt to 'solve' explainability would likely be an extreme reach-goal that should not be undertaken lightly. That being said, given that model explainability is a main issue for AI-ML we should practice due diligence and ask: How might an object-relationship-space framework help or not help or address or not address issues of explainability?

One way in which having an Object-Relationship-Space framework could help with various 'model explainability' issues, is not directly with a particular startical explanation calculation, but with what you might think of as a secondary set of problems that are very much a part of people's concerns about 'explainability,': a lack of clear definitions. By clarifying and disambiguating and framing needs with projects and specific users, making sure we have as clear an idea as possible about what people are really needing and asking for may go a long way towards solving at least so human needs regarding AI-ML Model explainability.

Another way, building off the first, is that in some or all these areas, specific tests may be designed to gather more information about the issue.

Some ways in which definitions can be clarified:

1. To better map out the problem space and create an 'Explainability, Reliability, and Security' problem space.
2. To help disambiguate specific issues, including:
 - identifying the correct type of problem or need:
 - accountability
 - redundancy
 - training data diversity
 - sub-component object-ability mismatch
 - understanding details of model input and output
 - system failure
 - finding double-standards
 - excluding or redirecting issues that not related to 'Explainability, Reliability, and Security'
3. To identify and define undefined or misunderstood elements.
4. To help define project specific needs.
5. To help define practical solutions that meet well defined needs.
6. Identifying model analysis methods that may meet the user's needs.
7. To match the right kind of resources to the specific kind of need:
 - emotional
 - accountability
 - security
 - system failure
 - system inconsistency
 - system monitoring
 - system reporting
 - redundancy
 - analysis
 - etc.

8. Standards for Concerns, Claims, and Assistance Requests

- A Telser Rule type problem, where a project participant or user keeps moving the goal post because they do not know, and have not defined, what they want.

Example Concerns, Claims, and Assistance Requests

- disambiguation of request
 - "I am concerned but I don't know what about exactly."
- reliability: adversarial
 - "I am concerned that X-user will feed adversarial inputs into the machine."
- reliability: redundancy
 - "I am concerned about what happens to users if the whole system crashes."
- reliability: full stack failures
 - "I am concerned that low level power failures and system crashes are not being explained or factored-in to how output is handled." (Like Big-Blue chess)
- reliability: Non-transference of ability between components
 - "I am concerned that not all components in the whole system have the same project-object handling ability. NLP is great, images are bad, I use both together."
- feedback for future corrections:
 - "I am concerned about mistakes being able to be corrected."
(e.g. ChatGPT reportedly uses reinforcement learning from human identified mistakes)
- Reliability: areas of insufficient-training ('hidden women')
 - "I am concerned about representation in the training data." See book ('Hidden Women')
- Model & Feature Analysis:
 - "I want to see a representation of the model for this answer."
 - "I want to see a back-track for this output."
 - "I want to see what training data were used for these features."
 - "I want to see the confidence levels of these different outputs."

More Disambiguation Examples:

Reasonable, unreasonable and undefined request examples:

- How can this system remove my accountability? (unreasonable/bad)
- Vague peace of mind. (undefined)
- Not disrupting use and users (reasonable)

- Fulfilling an ideological demand (unreasonable/bad)
- Flip-Flopping ever-changing demands (unreasonable/bad)
- fraud-gang use (unreasonable)
- wanting to create problems (unreasonable)
- wanting to stop use of data (unreasonable)
- wanting 'convenience' (unreasonable)
- wanting 'simplicity' (unreasonable)
- wanting instant solution (unreasonable)
- wanting violation of schedule tautologies (retroactive requests, etc.) (unreasonable)
- wanting passiveness (unreasonable)
- wanting comfortness (unreasonable)
- wanting a potemkin village (unreasonable)
- wanting to obfuscate project framework (unreasonable)

Real Needs but Other Types of Needs:

Satisfying people's anxiety is a real need, and likely it can be done, but the person's problem and need likely should be disambiguated. If someone has an emotional need, that is a real need, but in most cases a person's emotional need will not be met by talking with them about disproving their null hypothesis. A wider set of resources may need to be available to meet such needs.

Defining Tests

As well as defining needs and questions, specifying the handling of objects may also help with designing tests to evaluation how a system is performing:

- areas for reliability testing
- areas for security testing
- areas for component-interaction within the AI system (e.g. where not all components have the space object handling abilities.)
- training data representativeness tests
- embedding space connection testing

History:

The history of science and statistical explainability and the social-cultural history of scientific explanations is not only a huge topic but a massive can of worms. Tools for scientific testing are much newer than people think. The tension between pro and anti science camps has been ferocious. Disputes between different testing approaches have been ferocious. The history and language of the topic is convoluted and interdisciplinary.

Even without the topic of AI-ML (Artificial Intelligence and Machine Learning), the topic of scientific-explanations is a huge controversial topic over which even professional scientists are reduced to shouting matches. Over time we have made and will make progress, but this is not a clear cut area where everyone agrees on the same history and terminology and domains.

In many cases there is a popular misconception that engineering happens via a kind of pipeline or conveyor belt that starts with 'absolute scientific proof of explanation.' Two possible examples, pharmaceutical drug-approval-applications and boat/ship design. In real life, the emphasis is on exhaustive reliability testing, because we don't know "why" or "how" exactly a good design works or a bad design fails. But we can through exhaustive testing see how reliable a product is and put safeguards and redundancy in place to make it better. Even just on the level of having people agree on terms and concepts, trying to explain to people that the confused concept of 'scientific proof' is a misconception that does not exist at all does nothing to dissuade the person from using the concept, and demanding to get something that doesn't exist. And it does not help that Journalists frequently fuel misconceptions about the nature of STEM.

For example, 'explanation' has in many cases been institutionally defined as a cartesian correlation graph, which of course does not 'explain' anything (let alone "prove" anything). Yet people cling superstitiously to this or that practice without clear communication about what is happening.

Model Explanation is an important area where I am confident that we can make progress, but popular ill-will, misunderstandings, and general lack of education present often insurmountable social obstacles on top of what are already significant technical difficulties. That being said, if air travel has been made as safe as it is, if hard-drive recovery has been made as good as it has, if we can send people to mars and mine asteroids and design genomes and all the other things we do, then provided we can calm down, breath, and play nice, we can work out the problems needs and solutions for 'Explainability, Reliability, and Security.'

2.8 **Mind-Space:** Mapping a general problem-space (or spaces) for AI & Mind

Universality:

A common question in science fiction and the real science of astrobiology is how different we (h.sapiens-humans) may be from radio-civilization-type (Drake Equation) organisms on other planets in the universe? Will they think in anything like the way we do? Will it be possible to communicate with them at all? Will we have anything in common with them? What is local and what is 'universal'? Will they have the Portuguese language on their planet? Probably not. Can they have binary boolean logic? Yes, they can. Will they have base-10 number? They may not, but they could. Could they convert whatever number base they use to bases 2, 10, 16 etc. as we use on earth? Yes, conversion is possible. Will they use sound to communicate? Maybe not. Will they see in the same visible light spectrum? Maybe not. Will they experience time flowing in the same direction? Probably, yes. Will they dance? Maybe. Will they communicate? Probably. Will they have games? Probably. Will they play chess? Probably not. Will they experience the same phases of development, participation, and decline, as humans do? Probably some of them to some degree.

Out of the space of what is possible, what is more or less universal? So we can pick a few very conservative choices, like boolean logic, and (depending on who you talk to) much of mathematics. But what else?

I would argue that the same intersecting/interlocking areas that we have been using here are also tautologically (by definition) universal:

- STEM
- Projects
- Participation
- System Collapse
- Categories of Types of Systems
- Clear Definitions vs. Disinformation
- etc.

(And possibly the object relationship spaces for project-objects may also be universally share-able.)

Defining "mind"

In our context of interlocking & intersecting features (STEM, projects, participation, clear definitions, etc.): if we can assign a role to X_system as a participant in the project, where X_system will

- interact with other participants,
- have responsibilities including communication and observation,
- perform tasks,
- not exacerbate system collapse that disrupts the project,
- etc.

if these conditions are met, then 'mind' and 'mind-spaces' are appropriate terms to use to describe X_system, whatever X_system is (a homo sapiens-human, maybe a seeing-eye-dog, an AI robot, etc.).

Q: Does this mean that participation in a project is the only kind of 'mind'?

A: No. There will be plenty of debate about the consciousness of something which is incapable of carrying out roles in a project (whether that something is a cat, a jellyfish, or a remarkably irresponsible human).

Q: Is project participation sufficiently inclusive to overlap with 'mind' for a clear and functional definition that covers handling and managing a spectrum of case-by-case locally defined project-objects(in object-relationship-space)?

A: Maybe.

Participation space: "Participant" vs. "Person"

There are a number of parts of participation as relates to projects and intersecting areas (STEM, projects, participation, clear definitions, etc.). One slice is that as of 2023 h.sapiens globally are following the same however-possibly-apocryphal pattern as local words for 'person' referring only to members of that local tribe/gang or clique. Either as a subset of project-space or as part of self-identification space or drake-equation

mind-classification, h.sapiens need to do a better job of generalizing a 'participant' so that it includes various important areas including what kinds of things may be participating.

There are two important categories of items for generalizing participation:

1. Biological, Machine, ET
2. Pre-Participant, Participant, Post-Participant

Each group in the first category (1. Biological, Machine, ET) contains many sub-categories among which there can be any type of recombinant combination:
e.g.

Imagine the following chimera hybrid: human + horse + crow +
large_language_model + photosynthesizing_algae + self_driving_car + ET-octapus

Then on top of that you have the status of whether the individual is still-developing (a child) pre-participant, an 'adult' (participant), or a living or deceased elder (post-participant).

- Developmental mindspace and the dynamics of learning (pre-participant)
- participant functions, capable of full responsibility in projects.
- Views diverge on characterizing post-participants.

Part of what characterizes discussions of 'mind-space' is a generalization or abstraction that crosses over or make-combinations in ways that are normally not done. Not all possible examples of this will be deemed relevant to all h.sapiens-people but hopefully some will.

Example:

A social task is accomplished by 3 teams:

- Team 1. h.sapens-humans
- Team 2. AI-robots
- Team 3. ET Aliens

All have very different physiologies, yet appear to think about and solve the problem in very similar ways. What is similar or different about the 'mind-space' of the three groups?

Example:

A learning task is accomplished by 3 teams of supervised and still developing very young individuals:

- Team 1. h.sapens-humans
- Team 2. AI-robots
- Team 3. ET Aliens

All have very different physiologies, yet appear to learn and develop in some similar ways, including inabilities and obstacles to development. What is similar or different about the 'mind-space' of the three groups?

Example:

A performance task is accomplished by 3 teams of injured individuals who need rehabilitation to relearn skills after serious injury:

Team 1. h.sapens-humans

Team 2. AI-robots

Team 3. ET Aliens

All have very different physiologies, yet adapt and recover in similar ways, including inabilities and obstacles to recovery. What is similar or different about the 'mind-space' of the three groups?

Example:

A discipline task is accomplished by 3 teams of individuals who need to work out a problem involving disruptive behavior:

Team 1. h.sapens-humans

Team 2. AI-robots

Team 3. ET Aliens

All have very different physiologies, yet behave in similar ways, including potential destructive patterns or equilibria. What is similar or different about the 'mind-space' of the three groups?

Signal Coordination & Problem Solving:

Another perhaps more abstract but in other ways still very concrete part of 'mind space' are perhaps edge cases or specific cases of how various kinds of often not conscious in the drake-equation sense of the term species nevertheless solve problems with information and signals.

e.g.

A h.sapiens-human

An agile-team

A wolf

A wolf pack

An ant

An ant colony

A tree

A jellyfish

A slime-mold-colony

A server

A network of servers

A mushroom

A combination of plant species sharing information about pollinators and herbivores.

Note: This need to generalize terms for participants and projects and to more clearly defined terms such as 'person' is largely why I awkwardly refer to "h.sapiens-humans" in this paper: because this paper is specifically about a group of potential project-participants that includes h.sapiens-humans, AI and

ET Extraterrestrials, including all manor of hybrid combinations thereof, including other bio-tech additions, so the term 'person' and even 'human' are at least in this paper arguably ambiguous. And the goal here is to be as clear as possible in defining and using terms.

As we map out universal spaces, we may traverse our familiar interconnected tree of structures to see what areas of spaces may branch off from there. (Is there a particularly good or bad or non-arbitrary way to arrange these?)

- project space(s) :

- system productivity space(s)
- participant space(s)
- object relationship space(s)
- system status and productivity space(s)
- scout-values space(s)

- definition spaces(s)

- object relationship space(s)
- definition behavior space(s)
- general STEM space(s)
- general system collapse space(s)
- categories of types of systems(s)

system status and productivity space(s) :

- system disorder spaces(s)
- system recovery spaces(s)
- disturbance regime spaces(s)
- system fitness space(s)
- scout-values space(s)

(It is easy to forget how little we know about our own mind-brains.)

Learning, Perception, & memory Space(s) :

- training, and development spaces:
- non-transferring skills
- non-automatic learning
- automatic vs. non-automatic
- hard-coded learning
- adaptable learning
- transference vs. non-transference
- reversible vs. non-reversible learning
 - unlearning a problem
- static vs. use-it-or-lose-it learning
- memory-medium issues
- short term vs. long term perception
- short term vs. long term memory
- memory-data sharing and interfaces

('I can remember it for you wholesale')
- wire-together-fire-together issues
- concepts vs. knowledge-facts (and gray areas)

Potentially Non-Universal Patterns

Between h.sapiens, AI, and ET, there are a number of 'patterns' that we do not have confidence about the universality of, and herein may lie both an interesting and potentially practical frontier of discovery and exploration.

Note, some of these may cause of controversy between h.sapiens to be discussed:

- language
- Normal Doidge's network neuro-plasticity repair functions
- Normal Doidge's network wire-together-fire-together training
- Normal Doidge's network wire-together-fire-together disorders
- Normal Doidge's general 'Noisy-Network' disorders
- mindfulness
- reacting to events before they have been selected to happen
- Khanamhan Tversky system1 system2
- dreaming
- remote viewing
- the default mode network
- suspension of the default mode network

The Matrix

From quantum information theory to artificial neural networks, higher dimensional tensors, arrays, matrices (and other words that people from different disciplines can fight over the exact definitions and uses of) seem to be full of discoveries, surprises, and mysteries.

The whole nature of how higher dimensional and lower dimensional patterns and data exist and interact is, at least as of 2023, still an area of speculation. What is this mysterious space? What other forms and abilities will come from it? How else can it be used in engineered systems? (How will quantum information theory shape the future of machine learning?)

Was Imanuel Kant correct when he speculated that what we call reality is a perceptual interplay between higher dimensional data structures that we cannot comprehend, and lower dimensional data structures that we pull into existence and attach our notions of the world to?

Which is the 'primary' fabric of reality, the higher dimensional data? The lower dimensional data? The interplay between the two?

'Acts of perception' that slice 'latent manifolds' through higher dimensional information space?

What are the limits of the deep learning networks that people have cobbled together?

Are there inherent tradeoffs between 'generalization' and project-specific deployments?

What will happen to the human mind if and when we hybridize the h.sapiens brain together with AI-ML technology?

What happens when the default mode network is shut down?

How does remote viewing work?

How do people react to questions before the questions are selected?

What is the topology of mindsphere?

2.9 Object Relationship Based Testing

- ethics testing
 - system collapse
 - system fitness
 - system epidemiology
 - project based scout values
 - system & definition behavior framework
- participation testing
- object handling testing
 - Kasparov Event Horizon testing
 - project-space objects
 - schedules
- AI Component Tests
- AI Architecture Tests
- AI OS Tests

Part 3: Discussing the discussion of AI:

(Review)

AI must be able to handle **objects** in the following inter-related contexts:

1. object-relationship-spaces
2. project-object-database (in project-framework)
3. project-participants (in project-framework)

Such that these are defined as existing outside of the AI, and that so long as the AI effectively deals with these, it does not matter how the AI does so.

3.1 Definitions of Terms

"Define your terms,
or you and I
shall never understand one another."
~ Voltaire, *Dictionnaire philosophique*

3.1.1 Terminology Issues 1: The tangled Semantics of h.sapien-human ability.

3.1.1.1 The Auto-General-Transfer-Instant-Human Framework Hypothesis

Somehow a 'standard paradigm' has congealed in the AI literature regarding

- A. what AI cannot do and why, and
- B. what h.sapiens humans can do and why.

This is not so much something that people set out to put on a firm foundation of experimental, repeatable, falsifiable, results, but a framework or theory or hypothesis that people refer to because they appear to perceive that everyone else does so and that it is just true to they should also refer to this set of explanations as simply being true. I object to this.

This paradigm (in the Kuhnian sense) is shared by all the AI books I have found. I want to briefly and clearly, and if possible not-rudely, explain how it is flawed.

Auto-General-Transfer-Human Framework Hypothesis exaggerates h.sapiens-humans abilities and uses a circular logic to 'explain' that AI can't do what h.sapiens-humans do because AI does not have mysterious undefinable abilities that h.sapiens-humans don't actually have.

h.sapiens-humans understanding-intelligence-consciousness =
generalization,
transference (to new uses, contexts,
novel situations, other skills, etc.),
automatic learning,
instant learning

Possible Context:

1. AI Researchers have a legitimate goal in avoiding the past mistake of over-promising strong-AI which in the past led to 'AI-Winter' periods where funding and support significantly fell, leading to long term damage to the academic foundation of research. E.g. Can you think of academic research from Academia in more recent than Hinton in the 1990's? And it is described as

miraculous that he 'hung in there' despite non-stop intense pressure against him.

2. A main goal is to argue against alarmist clickbait journalism that repeatedly falsely reports exaggerated claims about AI 'surpassing' human abilities. Most books dedicate sections to debunking such headlines.

3. I'm not sure if this is a satisfying excuse on their behalf (just an idea I had, my fault if it is rubbish), but 'science of mind and consciousness and learning' is likewise a 'career limiting decision' and so there is no robust academic science of consciousness and learning to use.

4. It is very difficult to avoid an echo-chamber where everyone starts to echo the same ideas regardless of well founded feedback.

I do not mean to pick on Francois Chollet or attack him personally. I think his is fabulous, his book is fabulous, and I think he did the best job of explaining this paradigm so I quote him, I do not mean to hold him personally responsible for having originated it nor should anyone else.

A wonderful encapsulation of the auto-general-transfer-human framework hypothesis comes from Francois Chollet's Deep Learning with Python 2nd edition, in brief in chapter 5, and then much of chapter 14 elaborates:

FC DLwP 2nd-ed ch5 pp130

Interpolation can only help you make sense of things that are very close to what you've seen before: it enables local generalization. But remarkably, humans deal with extreme novelty all the time, and they do just fine. You don't need to be trained in advance on countless examples of every situation you'll ever have to encounter. Every single one of your days is different from any day you've experienced before, and different from any day experienced by anyone since the dawn of humanity. You can switch between spending a week in NYC, a week in Shanghai, and a week in Bangalore without requiring thousands of lifetimes of learning and rehearsal for each city.

Humans are capable of extreme generalization, which is enabled by cognitive mechanisms other than interpolation: abstraction, symbolic models of the world, reasoning, logic, common sense, innate priors about the world--what we generally call reason, as opposed to intuition and pattern recognition. The latter are largely interpolative in nature, but the former isn't. Both are essential to intelligence. We'll talk more about this in chapter 14.

h.sapiens do not:

- learn everything automatically
- learn instantly
- learn quickly
- transfer learning to novel situations
- transfer skills to other skill-areas and applications
- correct mistakes in past learning
- generalize automatically from anything to everything else

- manage novelty
- do 'just fine' with extremely alien, 'other,' novel experiences
-

h.sapiens do:

- reject and attack schedules
- reject STEM
- reject project management
- reject and attack data
- rapidly forget
- destroy themselves violently
- destroy each other violently
- destroy projects violently
- destroy STEM violently
- attack ethics in principle
- attack best practice in principle
- deliberately engage in counterproductive fraud and corruption
- shoot the messenger
- construct potemkin villages
- blame victims
- bully
- torture

Terms such as 'reason' 'symbols' and 'common sense' are not clearly defined, but the situation is worse than just that. These are very problematic terms that have no clear agreed upon definition, have immense historical baggage (see: western misogynist apocalyptic teleological eschatology, there's little baggage to start with), may be entirely fictional reifications i.e. may not exist at all any more than aether or phlogiston or notions in the past which were failed attempts to understand how the world works.

Perhaps the main reason why, years before ChatGPT I started putting together the Object-Relationship-Space framework, was that the auto-general-transfer model was so disconnected from the details of actual AI projects.

From what I can see, one of the main reasons why so few people saw Large Language Models' ability to handle objects (which generally isn't articulated anyway) is that everyone in echo-chamber fashion convince themselves of the auto-general-transfer paradigm, that somehow there was a special-sauce that all h.sapiens-humans have that allows them to use automatic-general-transfer-ness pixie-dust to learn and transfer and generalized everything automatically instantly, that this is what understanding-intelligence-consciousness is, and that AI simply doesn't have this pixy dust sauce so AI will never be able to be 'understand.' The lack of detail and granularity in this so-called explanation is astounding.

This has led to truly astounding statements about the human ability to learn and understand which seriously makes you wonder if the author has ever met or observed a human being.

- children learn to cross the road safely without ever being in danger of being hit by a car
- there is no culture shock

Not to mention that this completely flies in the face of the vast majority of human history:

- life for most people for most of history has been extremely uniform
- human learning and progress is very slow and tenuous
- people who threaten daily-familiarity are outright killed
- 'the other' or anything not familiar, is killed
- the words 'foreign' and 'alien' in English are pejorative terms
- parents and teachers are frantically paranoid about exposing children to anything 'unfamiliar'
- learning in general is excruciating and people simply hate it
- periods of 'disruption' of norms result in complete mental and social explosion into maladaptive disorder
-

3.1.1.2 What is the STEM evidence about the underlying mechanisms supporting how people think?

There are multiple compounding problems here:

- We do not know.
- We do not know that we do not know.
- We do not communicate clearly about what we know and do not know.
- We do not know that we do not communicate clearly.
- We do not know if we are asking the right questions.
- We do not know enough to know if our questions make sense yet.
- We do not know that we may not know enough to ask the right questions.

We have speculations, aspirations, and lots of disagreements. You've got to start somewhere, so let's just start, but it's a mess.

There is no consciousness-ology.

There is no education-ology.

3.1.1.3 How does it matter, how does this apply to AI discussion and design?

One hopefully pragmatic approach is that we should try to be clear about what is speculation and what is analogy, vs. what we just do not know about in detail yet and perhaps ever.

Using an imperfect analogy to explain something well enough to get a general idea across is probably fine, maybe great. Taking an imperfect analogy

literally, and or using circular explanations where several undefined terms all define each other, is going to create liabilities and problems for future-us.

Whether it is aerodynamics, or ship-building, or pharmacology, or AI-ML, we can use best practice and STEM and empiricism to test and build solutions that testably work to solve well defined problems even when we are either wrong or in the dark about exactly how those solutions work. It is good to find working solutions. It is good to improve our understanding of how things work. Those two are not the same, do not need to be the same, and do not require each other in order to exist.

Terms such as 'generalization' have multiple definitions and uses and meanings, and should be used and maintained pragmatically (not allowed to become problematically confusing).

Generalization, h.sapien-mind, and general problem-space for mind-space (for AI and biology) are important and interrelated areas, where a serious problem with defining h.sapiens can lead to problems with understanding generalization and general mind-space.

(And since it very absolutely predicted that ChatGPT would never happen, it will be interesting to see what evolves next in what how these same people expound this ideology.)

See:

- section on generalization:
 - 3.1.18 Generalization
- section on general mind space:
 - 2.8 Mind-Space: Mapping a general problem-space (or spaces) for AI & Mind

3.1.2 Terminology Issues 2: What has been defined or is not-defined?

3.1.2.1 A simple 'Can you explain?' test:

- The 'Can you explain what you are talking about?' rule:

If you are reading or discussing and the person cannot or will not explain what they mean by an inadequately defined term, then red-flag that term as undefined, prepare to ignore it, and move on.

This can be significant and whole discussions may end up being useless.

Confronting the undefined:

What should the reader do when they come across undefinable terms?

1. Try to use context for the reference to get a sense for how the author is using the term. If it is a transient sloppy reference that happens once in the book, take a guess (at what you think they should have said) and shrug it off. But if the term is central to an argument they are struggling to make as they recycle various confused terms...that is a bad sign.

One example of a consequence of an undefined goal or standard to test, is what has been called *Larry Tesler's AI rule* (though it is sometimes attributed or misattributed to other people, as often happen with quotes, especially very poignant ones like this one)

"Intelligence is whatever machines haven't done yet".

https://www.nomodes.com/Larry_Tesler_Consulting/Adages_and_Coinages.html

'moving the goal post' is another common phrase related to this pattern. If AI is defined or left undefined as a vague reification, the label can be arbitrarily given and revoked to anything, perhaps connecting to Karl Popper's notion that an untestable model that can be used to explain (or rationalize) any outcome is not useful.

Part of what I think is very significant about "Large Language Models" such as ChatGPT, is that it can be given concrete tests which it often passes based on object relationship space framework contexts, meanwhile Tesler type discussions flood the internet where people gesticulate in any and all undefined directions arbitrarily, which, as Karl Popper would say, is not accomplishing anything.

Example of terminology issue:

in section 6.2.4, page 164 of Francois Chollet's Deep Learning with Python, he says:

"...a logistic regression model has statistical power on MNIST but wouldn't be sufficient to solve the problem well. Remember that the universal tension in machine learning is between optimization and generalization. The ideal model is one that stands right at the **border** between underfitting and overfitting, between undercapacity and overcapacity. To figure out where this border lies, first you must cross it."

It is difficult to know how exactly to interpret these terms. Are these jargon terms for math abstractions? Are they rule-of-thumb terms for experienced model trainers (and Francois Chollet has as much experience as anyone)? Are they colorful analogy terms for popular readers? Or are they terms that people have come to use by group habit and no one really knows why they use them?

Optimization vs. generalization

undercapacity vs. overcapacity

underfitting vs. overfitting

(which Melanie Mitchell Called: underfitting vs. overgeneralization)

Francois Chollet said "the universal tension in machine learning is between optimization and generalization" where generalization is what he repeatedly explains is the goal.

Then he says: "The ideal model is one that stands right at the border between underfitting and overfitting, between undercapacity and overcapacity." where overfitting is NOT the goal.

As stated above, how are we supposed to interpret these terms? "Optimize" vs. "generalize" vs. "fit" vs. "capacity" and this is coming from someone who ends the books saying with absolute certainty that large language models will never succeed in doing exactly what OpenAI's large language model GPT did only a few months after the book was published. So how are we supposed to interpret this top of his field, literally the author of the standard software,'s view which produces basic statements about how the models work which are profoundly wrong?

Yes, nitpicking about editing aside, these are dichotomous concepts and the idea is that the process of model training aims to move the mode between two types of not-working into a third 'works-now' state. Do these terms actually mean anything? Do we have any idea what is actually going on? Do these terms contain meaningful information about what is happening?

3.1.2.2 Context and use:

3.1.2.2.1 If the term is being used as a temporary place-holder because the context is trying to arrive at a definition, then not having a definition at the beginning is not a problem. But if there is a presumption, assumption, declaration, etc., that the term has been defined when it has not, so that there is no attempt made or intent to in the future ever define the term, that can be a major liability.

3.1.2.2.2 If the term is a one-off not related to main topics, merely an aside in fuzzy detail, it can likely be safely ignored.

3.1.2.2.3 If the use of the term is a writing-practice failure (an error or oversight by accident of whatever various type) and by context a more clear term can be substituted without ambiguity, do so and move on.

3.1.3 Terminology Issues 3: Navigate Jargon Pragmatically:

- 'Artificial Intelligence' is not so bad
- "Bias"
- "Generalization"
- "complexity"

A problem which I often associate with the 'Tomato is not a vegetable' mania that grips at least the continental US with surprising ferocity, is the confusion of a technical-jargon term in one context with something else, for example a common-use word, another jargon term, another context, etc.

E.g.

The term 'generalize' is often a, often the, technical Jargon term to describe the 3rd option between "Overfitting" and "underfitting" (why they don't call it 'good-fitting' is beyond me).

Regression is perhaps a classic example of a jargon term which has become the official STEM technical jargon for a technical meaning, despite the fact that there is no logical connection between the 'normal language' meaning of the word and the technical-jargon meaning. In this can you can easily see how this 'mistake' (or bad naming choice) happened: Sir. Francis Gaulton was looking for mathematical ways to analyze how genetic expression 'regressed to the mean' over passing generations, so this mathematical model approach came to be called 'regression.' However in the current form of that math-modeling-approach, there is no 'regressing,' so the name is perpetually confusing. (Perhaps like calling a door-stop a 'stapler' because you invented the doorstop while trying to keep papers together.)

There are various kind of 'jargon management' problems that come up when reading books about AI. 'Bias' is defined in different technical and non-technical ways, very confusingly. 'Generalize' is especially annoyingly defined in many different ways, and rarely does the author make an aside to explain their definition. Sometimes, if not often, 'generalize' is actually used as a technical term to mean the opposite of itself as a technical term...ooof. At least in computer-science, where naming-things correctly is a known challenge, you can improve the names (sometimes part of what is called 'refactoring' (which also gets defined in many different ways!)), but in earth-land, the 'rules of nomenclature' hold that once named forever will it be the same.

And relating to the reification issue, 'complexity' is an amazing term in the history of science, where it does have some rare (and totally different from each-other) jargon meanings, the common meaning has literally no actual meaning at all. So be very careful when you hear an author use 'complexity' and try not to imagine it refers to anything at all.

in Summary: "artificial intelligence '' is a jargon term, because it was used in a research proposal and program in 1956. It is not actually a bad term, but it is jargon. There is no implied logical connection between the real meaning of 'artificial' and the real meaning of 'intelligence.' So it is pointless to try to micro-analyze the component terms, like 'regression' (a term which in jargon has no meaning at all), 'AI'

3.1.4 Terminology Issues 4: Fictional frames of reference are bad:

A "baseline" is good. There are so many model evaluation methods in AI-ML that DS-AI-ML is essentially the forefront of STEM science on the nature of scientific analysis...yet for some reason people seem to have a blind-spot covering themselves:

- human generalization
- human instant learning
- human 'intelligence'
- human 'general-understanding'

- human automatic learning
- human automatic transference

3.1.5 Terminology Issues 5:

Problematic multiple meanings of unavoidable terms

"Abstract"
 "Complex"
 "Bias"
 "Dimension"
 "Matrix"
 "Tensor"
 "Parameter"
 "Hypothesis"
 "Symbol/Symbolic"

- generalization_1 vs. overfitting,
 and mystical generalization_2,
 "over-generalization" = overfitting (MM)
 concepts
- 'embedding vectors': higher level topic vectors & concepts

3.1.6 Terminology Issues 6: Beware Non-sequitur Conclusions

- Humans make terrible decisions and can't learn
 Therefore (find quote):

M Wooldridge, ~"humans are model for all AI"

- You need to add bias to models so they do generalize

Therefore: some models are bad because they are biased and no models are good because no models can generalize. ...what?

- terms that get used in different ways
- terms that different technical and colloquial definitions
- terms that have multiple different technical definitions:
 - bias
 - parameters
 - hyperparameters
 - dimensions

3.1.7 Discussions of Model "explanation"

- Double standard between symbolic and subsymbolic

3.1.8 Reification:

As a reader you should be aware of the more or less standard but not entirely common or easy term 'reification,' which refers to a situation where people usually mistakenly create a concept which they then imbue with concrete reality and treat it as though it exists, though it may either not exist at all or be a confusion of several things.

Perhaps one of the best examples which also illustrates how reifications get used socially is the children's notion of 'cooties.' On one level the 'cooties' game may be a play-gamification of the children's experience of being thrown into an unmonitored pool of disturbance regimes, but aside from looking at deeper significance: simply on the surface of things cooties do not exist at all in any way whatsoever, yet children pretend that some people 'have' cooties and then need to be socially excluded because they are in possession of something that does not exist (perhaps this is practice for later life when some people are arrested and ostracized for possessing plants with no published medical use which also by that legal-description do not exist).

Whether it is a scientific abstraction hoped to solve a problem like Aether or Phlogiston, or an initial geological 'catastrophe,' or something socially-bad like 'boroisee-ness' or 'jewish-chess', or something ideal and sought after like 'arieness' or 'high-iq-ness', or the amazingly still frequently mistaken conflation of scientific hypothesis testing with mathematical (e.g.) geometrical proof into a hybrid nonsense of 'scientific proof.', or simply 'cooties': reification is an easily observable phenomena of the human mind asserting reality, often with powerful belief and emotions, where there may be nothing whatsoever.

A classic example from AI history may be the notion of 'chess-strategy-mind-ness' which imagined a type of mental ability with specific characteristics which...does not exist or at least not as it was thought it. And chess may also be an example of a testing-issue. Perhaps the type of pattern-perception that a human chess-master uses is a specific kind of general human ability, but the test for that is not chess. This was part of the erroneous reified belief: that if a computer could play that that would absolutely mean the computer had 'passed the test' and possessed this 'chess-master-pattern-perception' ability. There are human chess-masters, and undeniably each of them has some set of abilities they use to play chess, but the single reified general-ability == passing-the-test notion was, as many reifications are, a complete illusion existing only in human fantasy which people cannot distinguish from reality.

Note: The term reification may be used in an affirmative way to refer to an abstraction or device which does have practical use. And perhaps the 'useful but easy to misinterpret' side is also something that may come up in AI and testing situations.

Phrases like 'solve AI' or 'solve intelligence.'

3.1.9 Definition Collapse: Maintain your definitions

- links:

https://github.com/lineality/definition_behavior_studies
https://github.com/lineality/definition_studies_draft

3.1.10 Potemkin Villages and Telepathy-Tests

Some things that may be in some ways similar to the reification illusion are Potemkin Villages and what I nickname 'Telepathy-Test.'

A 'Telepathy-Test' is my nickname for a test where a person (for example a teacher) uses an open-ended question which could be interpreted and or answers in a number of different valid ways, and considers 'correct' answers to be those that the teacher was thinking of and anything else is "wrong."

This comes up a few times in AI books where the author describes a 'test' question given to an AI, and proceeds to say the AI has failed if the AI did not pick one single path to solving the question, sometime explicitly (falsely) stated that there is no other possible interpretation

.

It is very interesting how humans who brag about their own 'theory of mind' abilities overlook their ability to fail to see things from another person's point of view.

A potemkin village is a sad and sadly real and even more sadly recurring historical phenomena of dictators being shown fake results of failed decisions and failed projects, because accurate reporting would lead to a lethal tantrum. Truth-cooties.

To less dramatic degrees this may be common in any institution that does not systematically use data to examine outcomes and make choices, and yet, data is usually loathed and shunned. How that is supposed to work out well I fail to see.

(email mitchel e.g. question)

Protesters:

The idea that this sentence must be interpreted in one way is not a clear test. E.g. here's one word of context that reverses the so-called one-obvious answer: Jewish. The protesters were jewish an a european town sometime in the past 1000 years: what is more likely, that that city allowed pogroms against jews or that violent jewish protesters tried to overthrow the city?

Non-General:

We may need to prepare ourselves for the idea that there is no such thing as general-understanding, in the same way that there is no real project with infinite scope, or no real data operation requiring infinite resources, no game with infinitely flexible rules,

That by definition, specific deployments and instances are as non-general as they can be.

A general database is not a good idea.

A general operating system is not a good idea.

A participation framework made only of general-individuals is not a good idea.

Mistaken Reification:

The term 'complex' often is problematically used to refer to a situation of total confusion where no one knows what is going on. There are various problems with this:

1. since the word has so many meanings, it is not clear if this is the intended one.
2. reification: instead of the term being seen as referring to a negative 'no understanding or form' people reify the term to refer to something affirmative with mystical properties, which in this case do not exist, so this causes a failure of communication and the equivalent of inviting someone to a meeting that won't happen.

"Complexity" and western teleology

Among the many varied and ever changing meanings and uses of the apparently cursed term "complexity" which perhaps as a general rule should be avoided at all cost,

there is a common use which may be seen to have roots in the standard western teleological model of the cosmos: that the universe is the gradual story of the emergent becoming of a super-great one-solution-to-everything alien higher-dimensional other-dimensional devine super-ness that is pulled into manifestation out of an inert dead evil bad-physical world which is benevolently destroyed in the process of divine manifestation. This is not only found in theological discussions, but also in discussions of 'pure science' up to the time of writing (2023) and no doubt will persist with western culture. e.g. the classic 1958 chess ai paper, a 2021 bristol university paper about the 1958 paper, Jon Hand's 20?? book surveying areas of current hard science.

It may be that this admittedly too absolute and simple model may be better than the opposite extreme of absolute nihilism and embrace of destructive disinformation. Ideally there are fruitful middle-ground with no problematic extremism.

3.1.11 Terminology and Interpretation of Intent:

- Azimov's Laws of Robotics
- ELIZA the Psychotherapy AI

In both the case of Isac Azimov's three laws of Robotics and Weizenbalm's ELIZA AI, there are completely perpendicular-opposite statements in the literature about the overall 'intent' and interpretation of these.

1. They exist to show a simple working solution: job done!
2. They exist to illustrate that there is no simple solution: job maybe never done!

These two overall narratives are about as different as can be, yet depending on who you read both "Asimov's Laws" and "ELIZA" can be one or the other. In my view a straight forward look at the original material

from the author strongly leans in the direction of option 2: a nuanced warning that cautions against simplistic solutions. But in the spirit of linguistics being descriptive not prescriptive, it is not proper or sustainable to try to control (especially retroactively) how language is used, and in that way these two concepts (just to pick two examples of this phenomena) have multiple contradictory meanings in the larger populations of dynamical real world language. And at the very least you should be aware that authors may use either meaning, and quite possibly follow the indeterminate-incompetence-and-malice model and waffle back and forth between multiple meanings. This happens, and as a reader you should be aware of how things word in the landscape you are trying to navigate.

Note: In part, this is often due to the sheer size of the AI topic. It is likely not possible for a human to read and understand everything and every historical foundation for everything, so I would caution against faulting a given author for not having infinite nuance and insight into everything. On the other hand, be on your guard for dangerous mistakes (be they caused by incompetence, malice, or indeterminate incompetence and malice, it does not matter (and may never be determinable) which).

3.1.12 Terms that people cannot define while pretending they can:

- Language
- Mind
- Explain
- Complex/Complexity
- Statistics
- etc.

3.1.13 local context specific definitions

3.1.14 Negative Definitions:

e.g. nul hypothesis
using system collapse & non-system collapse
to define policy etc.

3.1.15 indirect definitions & negative definitions

'fail to disprove the nul hypothesis'
value function & meaning

value, function & meaning as indirectly and locally defined as not
general system collapse = non-collapse

non-overfitting, non-underfitting

3.1.16 Participation

Modular-recombinant aspects of participation:
pre-participant
participant

post-participant
&
biological non-h.sapiens
h.sapiens
technological non-h.sapiens
ET non-h.sapiens

3.1.17 Generalization

The term generalization deserves its own corner of the universe.

Is model-generalization a jargon term or a misnomer (or a tragic combination).

There may be a circular logic in Francois Challet's explanation of models: model's fit a pattern by generalizing, and they generalize by fitting a pattern: how linguistically convenient.

Part of the question we need to ask now is: "How wrong are we?"

(not knowing what the received wisdom will be a century from now...assuming and hoping they are better informed than we are today, barring some dark age between then and now) how far off the mark is our very odd overall framework that we use to describe the target and function of AI?

The bogus narrative goes like this:

h.sapiens == person
people can generalize and machine's can't!
people have minds because...they can generalize!
machine's can't generalize...because they don't have minds!
And machines aren't people, because they don't have conscious minds.
And 'h.sapien' generalization is embedded in a hodgepodge of notions:
- h.sapiens learn instantly
- h.sapiens learn automatically
- h.sapiens transfer from any thing to anything
- h.sapiens can handle pure novelty by generalizing and transferring everything to everything
- h.sapiens know and remember and see everything.

So, "How wrong are we?" Are we just slightly wrong and we can push the definitions around a bit and phew, everything is fine again! Or is this a giant dumpster-fire yardsale splatter-painting of hot garbage and we have to completely throw out even concepts like 'generalization'?

This situation perhaps is a context for talking about a general problem-space of mind-issues or 'mind-space' as I tragically shorten it sometimes. We thought we had, we wanted to have, a general (see that word!) set of patterns across both human-mind-stuff and AI-pattern-stuff, and wanted 'generalization' to be the hub, the central gear in a beautiful cosmic clockwork (though I doubt clocks have a 'central gear').

THE problem is...h.sapiens don't do the generalization skills that the framework lays out...and they can't even generalize the concept of 'person-participant' which probably would have been a legitimate use of the term generalize. And quite possibly, 'generalization' is not the right term to use for machine learning does either, the part that happens in-between over-fitting (which is necessary, if counterintuitive) and underfitting (which is also necessary, you can't go straight to over-fitting). And 'fitting' might even be a misleading 'analogical description.'

We have a hodgepodge of terms and concepts that don't remotely add up, and no concept of the problem space they are supposed to help us navigate:

General vs. not-general
underfit vs. overfit
higher dimensional vs. lower dimensional
specific-instance vs. pattern
high-definition vs. low-definition (and why doesn't anyone talk about Herbert Marshall McLuhan anymore?)

Kant's refrain: reality is a perceptual interplay between higher and lower dimensional patterns.

When talking about model-function success, the term 'generalize' is a strange jumping-conglomerate of notions. We use the term to describe the information process, machine-math-behavior, the explanatory mechanism behind how and why the function works, the user-story real-world-use-case, and the human-desire-thought-intention behind deploying the tool, AND the human-equivalent brain process that does the same task (even though people and machine are not supposed to share any such thing) all at the same time. What is the likelihood that one word really covers all that arm-waving? Smells like garbage, sounds like garbage, looks like garbage...I don't know about you, but I'm betting that it is total garbage.

This raises the question how much we have any idea what process the AI is undertaking when it gets the right answer. I don't mean 'how it does it' I mean just what we're asking it to do.

Also, this term 'generalization' tends to be selectively illustrated with examples like picture classification, and the ability to classify and sub-classify. But how about NLP? For example, when an AI selects its own word-ish character-gram (made of letters, not human words) sets of patterns, do we still think of that as 'generalizing' to a nice clean (and possibly totally fake) category like an icecream flavor that h.sapiens invented? The AI is forming its own concepts but not using words...do we still want to call that generalizing to real-world-patterns, just like the master-man!

When we ask an AI a question like, say something about a political leader, or tell me what this book is about, is that correctly matching a pattern like a

number to a shape? When we ask an AI a Sally-Anne Task question about what a person in a situation thinks from their point of view, is that 'generalization' to an underlying pattern like, how the number seven looks?

"Latent" Pattern in 'manifold' hypothesis

Another strange mix of notions, is the idea that we don't know if the 'manifold hypothesis' of fitting to a 'latent' pattern somewhere in a higher dimension is true...and we also use that as an authoritative explanation for what is happening. This is kind of like saying: We'll no one knows if there is a Dr. Regulus Black working at this hospital...but you need to take your medicine because Dr. Regulus Black said so! ...Really? Because that kind of sounds like you have no idea what you're talking about and you're hoping I have severe amnesia.

3.1.18 Controversial Topics

When you are dealing with a controversial topic, is probably a good idea to lean into intersecting/interlocking areas: clear definitions

- Clear & Functional Definitions
- Generalized STEM
- Generalized Projects
- Generalized Participation
- etc.

3.2 What To Read:

Compare points of view:

One of the things that you may find right at the beginning when comparing the perspectives, advice, and wisdom of different notable authors in the field of AI, is that they do not say the same thing on many topics including, notably, what your relationship to AI should be. For example, in Michael Wooldredge's fantastically eloquent book, he opens by saying that his book is a conceptual discussion of a highly technical field and that you certainly can't expect to become practitioner of the AI technical arts by reading the book. But then in Franscoi Challet's (the person who created the Keras software package) Deep Learning in Python, he opens his book by saying that we are close to a time when anyone and everyone will have the tools to build and maintain their own AI, not just rare specialists, and encourages the use of Keras to democratize access to AI. Hobbson Cole in what is perhaps the best conceptual and technical book on AI (in my own view) may not even comment on your station in life, he just tells you to build an interactive AI and clearly steps you through how to do it.

This is just one example of a sub-topic of AI where different authors will present to you substantially different sets of possible and recommended options.

Another one of what I think (perhaps incorrectly) of as one of the best books about AI, is just a biography of Alan Turing written in the 1980's. "Enigma," which inadvertently, just telling the life story of Mr. Turing, lays out an unusually interdisciplinary narrative where computers and cybersecurity and chess and AI and pure-mathematics and statistics and the telecommunications industries and all very much a part of the same intertwined story; which is very different from the clean-separate-lines version of the history you get from people who are often experts in one field.

And in some ways people don't even mention the same cast of characters at all. George Bool and Douglass Hofstader and John Bayes and the inexplicably invisible yet essential Claud Shannon are sometimes missing entirely from a given version of events. (e.g. Claud Shannon was part of the group the created, organized, proposed, and ran the original Dartmouth AI research program that today's AI is still based on. For all we know he wrote the proposal which still defines the field, yet usually people don't mention him at all.)

'Possible Minds' is a great collection of essays about Norbert Wiener's 1940's 'Cybernetics' research movement which at the time was how AI was widely discussed, but that whole chapter on AI is generally not mentioned at all whatsoever in standard AI books.

And to some extent we are still struggling with the 1820's work of Ada Lovelace and Charles Babbage (also not mentioned, though in his day there were popular songs about Babbage's AI technologies), regarding the most simple questions of Q: what does it mean for a "machine" to use a "symbol." This perhaps simple sounding question will be deeply and open-endedly threading through much of this paper.

3.2.1 The Three-Legged Writing Stool

A Three Legged Writing Stool:

Leg 1: Tech Experts (Like engineers and scientists)

Leg 2: Writer-Communication Experts (Like Authors and filmmakers)

Leg 3: Research & Exploration Experts (like journalists and mountain climbers)

A Three Legged Writing Stool:

Leg 1: Tech Experts (Like engineers and scientists)

Leg 2: Writer Communication Experters (Like Authors and filmmakers)

Leg 3: Research & Exploration Experts (like journalists and mountain climbers)

The problem of the need for collaboration between writers and technologists, and dedicated-explorers:

Most writer's alone are not sufficiently expert in the domain knowledge (and when they write books alone, those books are unreadable un-edited jubbiliee of incoherent garbage that honestly don't even help other tech people); most technical experts are extremely bad at writing and ironically for being in STEM

are appalling bad at communicating in any way which is probably why STEM is not performing to spec; writers, obviously, are not experts in tech, so if they try to write about something they don't understand, they produce at much more read-able kind of useless garbage. And there's a kind of third leg of the stool sample here I think: unstoppable border-crossing explorers, the Edmund Hillaries and Magellins and people who don't just sit in the office re-reading the same books and re-discussing them with the same peer-writers. And this third leg may be the achilles heel of the h.sapiens species, as we seem to be terminally (literally 'terminally') preprogrammed to burn Giordano Bruno in public if he dares talk about anything unusual: either we stop doing this, we don't survive as a species because we've eviscerated our own senses and brains out of pure self destructive stupidly, which would be a very sad loss and failure.

And the explorers and maybe a tricky part, because they are neither communication experts nor domain experts, but we need them. And frankly they are a blessing that we don't deserve, as we're hell bent on destroying them because they produce something we need.

- The C Programming Language: Perhaps a classic of literature for all time. Brian Kernigan did such an inspired clear and poetic job of covering such a well selected set of topics in such a small read-able book. This book has probably had a disproportionately large impact on the history of technology, possibly being responsible for C being such a dominant language for so long. And ironically, that C The Programming Language has been eternally the only readable and useful book about programming may have lead to the festering of memory management problems which have become an international crisis, something I do not think we can in any way blame the author's for not psychically foreseeing and solving pre-emptively, given that aside from frantically now more than 50 years later trying to make a 2nd-Gen C language called "Rust" (tragically horrible name, wonderful wonderful language, it's our superhero saving the world literally).
- possible minds: Wonderful collection so you can compare writers with different background writing about exactly the same AI topic.

(See more specific comments in appendix.)

3.2.2 History

AI is a great place for a history-digging treasure hunt. The search will take you into through all kind of expected and unexpected areas and probably turn up quite a few surprises. Things that you thought were different but are the same, or thought were the same but are different. (Like cognitive psychology and artificial neurons.) Things that are older than you thought, things that were younger. And many items that are different in different books (or different parts of multifaceted histories).

AI is what should be a great coming together of education, biology, math, engineering, medicine, computer science, statistics, and more. But is it like the classic image of the rail-road project with worker standing around two pairs of tracks built to meet from opposite directions but that tragically/comically pass each-other in parallel rather than meeting?

The pinball-effect.

https://en.wikipedia.org/wiki/Claude_Shannon

is gold. Find out everything you can about what he said and did, but he is mysteriously not covered much if at all.

3.2.3 Interdisciplinary Area Recommendations

- Biology:
 - Ants
 - non-chordata: trees, fungi
 -
- Statistics
- Cybersecurity
- History of Science
- Philosophy of Science / Nature of Science
- Linguistics
- Non-linear Dynamics, Fractals, 'Chaos' & 'Complexity'
- Ethnobotany:
 - Rupert Sheldrake, Dean Raden: Maybe it's a dead end and a waste of time, maybe not. Reading a few short books by researchers about published research is how it's supposed to work. Science is not a popularity contest.

3.2.4 Do AI Projects

In case you have only read about AI, I highly recommend that you do some projects. There are many projects you can do, even on a mobile device like a phone or tablet, my using Google's Colab online jupyter notebooks.

Two books I recommend for projects are:

- Francois Chollet &
- Hobbson Cole

3.2.5 Book Recommendations

- short list: (all on audible)
- Melanie Metchel: AI
 - Michael Wooldridge: AI
 - Hobbson Lane: NLP In Action 1st edition

- Franscoi Challet: Deep Learning with Python 2nd ed

medium list:

- enigma
- the theory that would not die
- Significant Figures
- Melanie Mitchel: complexity
- possible minds
- we are all targets
- Postwar
- history of the future
- the signal and the noise
- cosmosapiens

longer list:

link

(See links in appendix)

3.2.6 Read classic Science Fiction: Back to Blade Runner again

It amazes me how people in 2023 who consider themselves true-blue science fiction lovers simply refuse outright to read 'golden age' science fiction from (very roughly) the 1930's through the 1960 (40's to 50's may be more orthodox). The difficulty of defining 'what is AI' rhymes in more than one way with the difficulty of defining 'Science Fiction.'

Theodore Sturgeon

Philip K Dick

Fredrich Pohl

(See appendix for expansions on topics from part 3.)

3.3 Examining Tests for AI: as discussed in the literature

(Under Construction)

3.3.1 - looking at winograd schemas

3.3.2 - Sally Anne Tasks

3.4 Empiricism & Influences on Model Architecture

In section 9.3.1 on page 251 of Deep Learning with Python Francois Chollet explains (here paraphrased and broken into smaller quotes to avoid the doom of copyright, I recommend that you buy and read the book):

There is perhaps a two-edged sword nature to Empiricism here, on the one hand Chollet blames an overwhelmingly empirical approach for a lack of understanding how models work, but at the same time his proposed solution is a very empirical approach of "ablation studies." Chollet describes "Ablation" as his preferred and recommended process of removing unnecessary parts from AI

that do not help testable functioning but rather are present by historical accident.

"Deep learning architectures are often more evolved than designed they were developed by repeatedly trying things and selecting what seemed to work...you can remove a few modules (or replace some trained features with random ones) with no loss of performance"

On Incentives & Purposes:

"by making a system more complex than necessary, [researchers] can make it appear more interesting or more novel, and thus increase their chances of getting a paper through the peer-review process. If you read lots of deep learning papers, you will notice that they're often optimized for peer review in both style and content in ways that actively hurt clarity of explanation and reliability of results."

"mathematics in deep learning papers is rarely used for clearly formalizing concepts or deriving non-obvious results- rather, it gets leveraged as a signal of seriousness, like an expensive suit on a salesman...The goal of research shouldn't be merely to publish..."

3.5 What do we do with Large Language Models & ChatGPT?

- Orthodox Tests ChatGPT can pass.
- Object Space Tests ChatGPT can Pass.
- How have various predictions fared, perhaps like chess, for what would have been needed to deal with objects?

- As in the conversation between a self described coder interviewer and the head of Open AI, there is no agreement or visible overlap in how they approach the topic of defining philosophical-intelligence. In this kind of environment where people are literally not talking about the same thing or using the same terms and the 'conversation' devolves into a group-monologue where each person drones on flailing with undefinable terms accomplishing nothing, if accomplishing nothing but having fun doing so in a coffee shop is the goal then that's is perfectly fine, maybe great art or something will come from the discussion., But if there are practical concerns and people are talking about specific system design issues, a framework such as Object Relationship Spaces may help to facilitate articulation and communication between people so that we are no talking past each other and perhaps even past ourselves where future-us won't be able to figure out what we were talking about because the definitions are so fuzzy.

- Sample Conversations in Appendix:
 - recommended: look at "Alien" film discussion
 - The general-inclusive vs. specific-deployment question.

If you are looking to make an on-edge deployment AI to solve a very specific issue with minimal resources, upkeep, attack surface, etc., is the question of whether or not chatGPT has some kind of general intelligence even relevant to your task? Given that AI-ML is a big-tent which includes a large

number of technologies, applications, even areas of math such as curve-fitting which are extremely general, how much of that will/should be by definition focused on a specific task or set of tasks, with no desired scope creep into additional user-features or system abilities/skills?

For those systems that would benefit from

In the world before chatGPT, we waved our arms and said "AI can't do anything, AI can never do anything." In the world after chatGPT we wave our arms and say "AI can do anything,." Both of these sentiments are not practical, productive, and well defined. Both of these are not intersecting with the key intersecting areas tools that we can use to leverage our abilities and understanding. Something big has happened, but we don't know what it is or what it isn't or what the next big thing might be, and we are not making our ancestors proud with our inarticulate bungling.

3.4.1 Communication and Describe Systems with a Well Defined Framework

3.4.2 The Tesler Rule Trap

3.4.1 The Empty Pronoun Loop

Part 4: Goals, Background & Future:

(In Summary)

4.1 Agenda & Goals

4.2 Background Concepts and Principles

4.3 Future Design Factors

Introduction to Part 4:

The goal here is to give a brief outline of areas that I recommend you look into and think about as part of thinking about AI. Ideally, this section will help you to expand and clarify the topic of AI, including my being frank about my Agenda (or at least what I am conscious of about my own agenda), background areas that are sometimes left unmentioned or are not clearly covered in AI books, and future topics to keep in mind for example to clarify areas where we want AI to do specific things.

4.1 Agenda & Goals

4.1.1 This paper's goals and agenda

(link to definition studies summary)

https://github.com/lineality/definition_behavior_studies

'We-can' statements.

The idea in this paper is that we can, and the agenda of this paper is that we should, connect the following areas (the same interconnecting/intersecting areas mentioned throughout this paper):

- general STEM
- general Projects & Project Management

- General System Collapse
- General System & Definition Behavior Studies
- General Categories of Types of Systems

to be able to define and navigate these areas

- Project-Defined 'boy scout' values
- Operationally Defined Ethics
- Machine-understandable Ethics
- System health
- System epidemiology
- System productivity

and apply that to this area:

- the area of AI Architecture & Operating Systems

Agenda, Goals Means Method Statement: "We-can" statement goals from Definition Behavior studies, a systematic study of general system collapse behaviors, which is one of the intersecting/interlocking areas along with STEM, Project-Context, Participation, etc.

We can succeed.

We can understand what's wrong.

We can make things work.

We can fix what is broken.

We can use non-automatically lost skills.

We can generalize STEM.

We can generalize system collapse.

We can use STEM to connect signals and reality.

We can connect STEM project management and ethics.

We can communicate, learn, and solve problems.

We can use "low-bar enlightenment."

(Taking 'potentially endless cycles of rebirth due to ignorance' as a metaphor/analogy for project-failure), in a context of projects not failing.)

1. The perception that perception can be fooled in principle and in practice.
 2. The perception that learning from failures does not happen automatically (and can, under bad circumstances, indefinitely not-happen).
 3. The perception that causality models can be wrong in principle and in practice.
 4. The perception that plans/goals can be incorrectly set (so that they fail to be achieved as set).
 5. The perception that repeating cycles of failure in practice result from errors in perception and planning (a proverbial 'wheel of samsara') without inevitable-automatic-learning based on that failure.
- ? (generalize categories of types of systems?)

- we can generalize system & definition collapse behaviors

- we can use categories of types of systems
 - we can use nonautomatic learning
 - we can find and fix errors in perception

 - we can organize projects
 - we can use clear descriptions
 - we can distinguish short term vs. long term
 - we can assign roles
 - we can check and verify
 - we can have policies on "errors and mistakes"

 - we can improve and cultivate perception by perceiving perception
 - we can prevent future problems
 - we can reverse damage from past problems
 - we can learn from the past
 - we can collect data
 - we can make and use policies
- (We can operationally define 'policy' as algorithms for non-collapse based on dynamics of collapse.)
- we can make strategies
 - we can make tests
 - we can make operational definitions
 - we can audit
 - we can publish
 - we can act with ethics, empathy and compassion
 - we can follow best practice

We can communicate:

- across space
- across time
- across cultures
- across generation-gaps & succession gaps
- across languages
- across types of participants
- across roles
- across projects
- across media of communication

- we can understand a spectrum of disinformation and clarification of information
- we can implement sustainable solutions
- we can fix what is broken
- we can prevent future problems
- we can reverse damage from past problems
- we can learn from the past
- we can collect data
- we can make policies
- we can make strategies
- we can make mandates

- we can make tests
- we can make functional and operational definitions
- we can succeed in a projects
- we can meet/deliver the needs of the target/user
- we can make progress

We can make progress by using information about the behavior of definitions: This is (system and) definition (behavior) studies, the field of study pertaining to the behavior of definitions.

These can-do statements are instrumentalist modular tool set areas. We can compliment instrumentalist modular tool set areas with narratives on principles and applications.

4.1.2 Defining your goals

try to follow intersecting/interlocking areas as a framework for a functional project space that your goals should be able to operate within.
(link list intersecting/interlocking areas)

4.2 Background Concepts and Principles

4.2.1 "Intersecting Areas":

From Abstract: "A repeating theme, context, and agenda in this paper is to pragmatically leverage the interconnected functionality of clear definitions, STEM, projects, participation, positive values, and productivity."

- interlocking areas: / intersecting areas:
 - Clear & Functional Definitions
 - Generalized STEM
 - Generalized Projects (project-context)
 - Generalized Participation
 - Generalized System Collapse
 - Generalized Categories of Types of Systems
 - Generalized Ethics, Duty & Responsibility
 - Generalized Definition-Clarification vs. Disinformation Attacks
 - (maybe) Generalized Definition Behaviors.
- value-function-meaning, boy scout-project-values,
- (maybe) Generalized (Sustainable) Productivity

(The overall goals is to bring together as many interlocking tools as possible over time, but exactly should be included now is unclear.)

Affirmative, practical, intersecting, well-defined, instrumentalist, non-ideological, testable, falsifiable,

4.2.2 input output measures...or next section

Input Output Measures are general but were developed in language education to specify curricula testing and IEP student performance measurement because h.sapiens-humans do NOT automatically-learn-transfer-generalize in these areas.

link to doc: https://github.com/lineality/input_output_measures

4.2.3 Higher Dimensional Frontier: Tensors & Matrices

4.2.4 Projects & Project Context

- STEM timeline
- concrete context for skills perception, etc.

4.2.5 Instrumentalism and Realism

4.2.6 Big Other Areas:

- Nonlinearity & "Complexity" Science
 - Definition Behavior Studies
 - General System Collapse
 - System Fitness
 - System "ethics"
 - Kahneman Tversky
 - automated coordination & decision making & voting
 - Project-Context Decision-Making
 - Involving Participants & Components
- https://github.com/lineality/Online_Voting_Using_One_Time_Pads

- General Learning
(project-context)
- *history of big tent of AI*
- *'complexity sciences'*
- *taboo areas of human mind studies*
- *Definition Behavior Studies & General System Collapse*
- *Generalized STEM*
- *Projects & STEM*

4.2.7 Gamification

summary of gamification:

My favorite way to think about gamification is to think about a scene (or two) from Star Wars Episode 4: A New Hope, when ships are giving the pilots, or gunners, graphical information about the target.

I will try to encapsulate the point with a loaded question: if you were in one two starwars ships in a dogfight, where one of the ships had an episode-4 type display for the gunner, and the other ship had a super high resolution 3d

physics engine first person gaming interface, with all the accompanying realities: extremely high energy cost, extreme sensitivity to network speed, constant glitching, constant lag, a giant bloated operating system that could stop to upgrade itself at any minute, pop-up applications that could bring down your targeting software, etc, etc. Hopefully the point is clear that your chances of surviving are higher if you have a super-simple system that could run on a microcontroller with little power, limited data, extremely fast, etc. Though it may seem 'counterintuitive' for there to be advantages to an 'old fashioned' system.

The idea of gamification in summary goes like this:

- a highly simplified problem-space reduced from the near infinite analogue data coming in through sensors has many advantages.
 1. one is that the reduced scenario can be handled using exponentially fewer computer resources (power, memory, processor speed, etc.) (which is the name of the game if you are a programmer: find the low-cost solution (or run out of resources trying...)).
 2. This reduced-scenario also opens more avenues for AI assistance or an AI solution.

(Maybe part two of that question: which ship would you want to be in if an AI was targeting? The one with the reduced task, or the one with a gazillion inputs and no clear objective? Or phrased differently, which would you prefer to bet your life on (if you had to pick one), the AI that learned to play the star wars Atari game at superhuman levels, or the AI-self-driving-car software that can't stop crashing into random objects? The gamified task is better.)

Assuming that I convinced you in that overly-brief summary of the value of a gamification reduction in the problem space to the simplest matching game, our real-world challenge is to teach an AI how to 'find the simple game' or gamify a large amount of incoming data.

(see appendix)

4.2.8 Ambiguous Equivalence:

- downside of analogies?
- higher
- manifolds
- symbolic subsymbolic
- system 1 system 2
- head heart
- right brain left brain
- parametric non-parametric

4.3 Future Design Factors

In addition to there being things in the future that we should be able to clearly describe, there are also background terms and concepts which will be important, for example work that we assume has been done and agreed-upon but which has not.

4.3.1 biology:

For any area of biology there are a number of general possible areas of relevance:

- 4.3.1.1** integration with biological systems
- 4.3.1.2** use or imitation (of biological functions)
- 4.3.1.3** compare and contrast for study and understanding
- 4.3.1.4** highlighting known areas of development and challenge
- 4.3.1.5** highlighting still not well understood areas of development and challenge.
- 4.3.1.6** highlighting predictable problems and pitfalls, challenges, etc.

While there are a number of biology areas that may or perhaps should come together, and in an alternate-timeline bio and nano tech developed sooner than AI, but on this earth in this timeline we may still have the goal (or perhaps the need but not yet the chosen goal) of merging AI into biology yet Large Language Model type AI (though that term may be obsolete soon, who knows) is being developed sooner and bio-tech is still largely embargoed by literal pre-renaissance religious fundamentalists, which is unfortunate.

Perhaps as an example of how a framework such as Objet Relationship spaces may be practical, in the above scenario it is somewhat of a moot or vague issue whether a completely undefined 'AI' consciousness intelligence body-less-robot has come into being or not, since there are no clear definitions and still from what I see most people (from what I see all people) still fall into the Tesler Rule and say, 'You have entertained me somewhat with your tiresome toys, but this surely isn't REAL AI.' and so long as AI is undefined any level of skill/ability could receive the same dismissive treatment. But if instead we set aside undefined terms, and instead take up well defined framework for skills and abilities and contexts such as project-context, we can now say (regardless of whether any who people will agree to apply any undefined terms such as true-intelligence-ness) that 'AI-ML' systems such as (so far we know, Large Language Model ChatGPT) can perform many basic object-relationship-space skills for general project participation. These project-participation-object-skills are not ambiguous. And it means that something other than h.sapiens can now participate in well defined projects as participants along with h.sapiens.

- 4.3.1.7** the science of sleep
- 4.3.1.8** the science of memory
- 4.3.1.9** non-chordata "intelligence" & decision making
- 4.3.1.10** science of mind
- 4.3.1.11** science of entheogens
- 4.3.1.12** science of mindfulness
- 4.3.1.13** Bio-Nano-Coded AI
 - currently: python or a systems-programming language, in future:
 - DNA and nano-tech in synthetic ~bacteria or cells
- 4.3.1.14** AI in synthetic organisms for terraforming planets, moons, etc.
- 4.3.1.15** DNA/RNA based digital information interface
- 4.3.1.16** the dragon project: Modularizing AI and Modularizing DNA

combined conditional expression hybrid genomes

4.3.2 cybersecurity and AI

Perhaps just as in the early days of software and networks, security and epidemiology, hygiene was not seen as being an issue or high priority, a future world in which AI systems are as much a part of the infrastructure as digital computers and networks may create a new 'attack surface' etc. where indeed there are security system-epidemiology issues, and perhaps we will have a better vocabulary for it. Though this past is prologue, we might even have a worse vocabulary for it and be hindered from understanding the problems due to our own self-mismanagement of language.

4.3.3 quantum information theory & under-the-hood optimizations

- sound coding and AI

4.3.4 "generalization" vs. deployment: managing development, production-deployment, and 'generalization' of project scope and liabilities.

4.3.5 Nanotech

4.3.6 Understanding Exponential Elbows

- The Fractal of Perception
- The Difficult to Predict Physical Events

4.3.7 'complexity' nonlinearity, dynamical, and systems sciences

4.3.8 Ethics, Projects, Best Practice & STEM

Two unhelpful directions that the discussions of 'Ethics and AI' go in are either:

1. The tragic mindset of nihilism that so many people seem to revel in. For whatever reason, people just love the idea, perhaps well symbolized by the current interpretation of the 'trolley problem' (regardless of that scenario's origins) where people are just determined to believe that nothing is possible, that all outcomes are horrible, that no one can agree on anything, that nothing can be done, and everything is arbitrary, in the ponderous celebration of nihilism to what end I really do not understand.

or

2. A nebulous sausage-making set of vague platitudes that are seriously lacking in clear definitions, and which therefore we just can't do anything with.

What I would very much like to do is steer away from these very attractive bad-equilibria, and help to steer people toward what I see as quite abundant low-hanging fruit in the area of, as this paper so frequently reiterates, intersection/interlocking areas of STEM, project-context, boy-scout values (or perhaps now just 'scout' values)

"Mindfulness" a Good Fit for AI (2023.04.05)

1. There is abundant low hanging fruit.

The term "Mindfulness" is a pretty good fit for talking about AI. While 'mindfulness' is likely viewed by some as being controversial or containing extraneous baggage, one of the long standing obstacles for AI is that just about all discussion of and research on mind and consciousness are seen by some outspoken groups as being objectionable. As outlined below, there are quite a few aspects of mindfulness broadly defined that fit squarely with the AI topics in this paper.

2. Terms and Fitness Concepts

Defining terms: AI, Ethics, Mindfulness

"AI"

This paper is focused on a project-context, and here "AI" refers to the big-tent of possible AI-ML areas as used in a project context (assuming best practice and intersection-areas for managing the project).

"Ethics"

Best Practice and Boy Scout Values in Projects with Interlocking Areas
(See sections on 'ethics' for more details.)

"Mindfulness"

Mindfulness is about awareness, in a way that can be very broadly broken down like this:

1. Awareness of Principles and Concepts of system-fitness (details below)
(System-Fitness Space in Abstraction)

2.1 Awareness of the local situation (where you are, what is happening around you) (Local Situation)

2.2 Awareness of the system-fitness of the situation around you. (Fitness of Local Situation)

3.1 Awareness of your context in the situation around you & what you are doing. (Your actions in local situation)

3.2 Awareness of the system-fitness of you and your actions in your situation. (Fitness of you and your actions in local situation)

These are all very relevant for completing the tasks assigned to AI components in projects.

Principles and Concepts of System-Fitness:

(signals, learning, communication, coordination,
*project-task-completion.)

- Degrading, Eroding, Weathering, Corrosion, of parts over time
- not-instant
- not automatic
- not transferring
- not general
- propensity to collapse
- propensity to obscure
- propensity for potemkin villages
- propensity for 'bias' (needs to be defined clearly case by case)

3. Areas from low hanging fruit to reach-goals and unknowns

- even the reach-areas are still squarely on-topic
 my personal agenda here is to systematize the practical,
 but it would be insanity to campaign against exploring the nature of mind
 within the topic of AI-mind-intelligence in a context of general
 mind-intelligence, that is would be about as self contradictory as you can get.

3.1 System-Fitness: Low Hanging fruit

3.2 Ethics: Requires interlocking/intersecting areas, but low hanging fruit

3.3 "Empathy/Compassion": ambiguous stretch goal, but the right topic

Participant-scale of projects = "Empathy/Compassion"

One possible concrete way of looking at 'machine empathy' in this context, is not so much a metaphysical reification of empathy, but an articulation of the assumptions of how project best practice 'ethics' extend beyond the narrow scope of the project details to the sustainable productivity of not only the participants (including stakeholders) and target-users, but also to other projects near and far in an eco-system of interlocking projects. It is completely consistent with a long-term set of goals and objectives that system fitness is valued and supported everywhere.

3.4 Default mode Network: very ambiguous, still broadly on topic

Other aspects of mindfulness research are still on-topic broadly as they regard the mind and the workings of the brain, which could in principle help with both new and better AI architects and with new and better integrations of AI into biological systems. However, the broader scopes of nature of mind, consciousness, and human neurobiology are probably beyond the scope of this particular paper which aims to focus on applying frameworks to AI in projects (though new discoveries which can happen any time may end up being directly applicable). At time of writing, the Default mode Network is a future topic.

To conclude, I am semantics-agnostic overall and my main goal is to push the topic and not get hung up on the term mindfulness and whatever baggage it might have or be assumed to have. Though given the lack of other terms that so squarely cover the relevant topics I feel a fair assessment is that the term is not at all a bad fit.

Note: "*project-task-completion": given that AI-ML can be used in possibly an infinite variety of ways for even more specific tasks, instead of attempting to list them all, here I simply say "tasks")

STEM, Ethics & Mindfulness: A Mindful AI Program:

- The new hard goal is empathy-compassion
- The Default Mode Network & Manifold Theory

4.3.9 Projects: Agile

- Agile as key historical development in STEM: project context
- Projects as main interlocking/intersecting area
- Machine Ethics from a Project Context
- Projects in AI System Architecture

4.3.10 AI and Code Testing:

- traditional code testing for AI deployment
- AI in code testing non-AI-deployment
- hybrid AI testing and projects

4.3.11 The Long Term Memory Storage Problem

- DNA
- quartz
- proteins (some outlasting DNA in fossils)
- 'fossilized memory structures'

4.3.12 The challenge of orientation and navigation in *mind-space*:

- avoiding collapse and contraction
- parkinsons & extension
- echo-chambers and silos
- non-automatic learning
- habit & atrophy: use it or lose it
- errors in wire-together-fire-together
- heterogeneity of equilibria

4.3.13 Human machine interactions, biology machine integration

4.3.14 Project-Context Decision-Making Involving Participants and Components

https://github.com/lineality/Online_Voting_Using_One_Time_Pads

Along with areas such as education, the mind, ethics, etc. yet another such strangely neglected blindspot area is a general area for hopefully can be described clearly enough as 'Project-Context Decision-Making.' Part of the problem is that not being 'generalized' perhaps in the same way that there still in 2023 there is no General-Drake-Equation-Person-Participant concept, only a 'our tribe are the true people, everyone else is a 'sub-human barbarian' (the british common meaning of 'foreign' and the Japanese literal meaning of 外人, according to Henry Kissenger the CCP's 'foreign barbarian' policy that has not changed since the bronze age, etc. etc.)

we only have a concept of: 'the local tribe member strong-men voting in government elections'

whereas we need not only a concept but a whole areas of technical sophistication around what can hopefully be clearly enough described as Project-Context Decision-Making. 'Voting in an election' are probably are not appropriate terms for all consensus and decision making in multi-participant projects,

This is a crucial area where most projects fail on this fail-before-you-start level, and in my experience schools are still actively moving in the opposite directly preventing any experience or skill development in Project-Context Decision-Making perhaps because of the unsightly chaos that results (because people have no skills) and that people have an aversion to what they hate: schedules organized projects accountability consequences, etc.

So the tragic and doomed status quo is a deplorable potemkin village in which the wagons are circled and anyone who talks about general Project-Context Decision-Making is jettisoned out and blamed for causing all problems.

So, we have a lot of work to do in this area to create not only concepts but infrastructure and time-tested systems that include not just h.sapiens humans but also various AI and subcomponents, etc. (plus other Drake-Equation participants, hybrids, etc.)

4.3.15 Question Space

4.3.16 self-awareness space

a shiny luxury or useful?

- chatpgt as an example:

does is matter what is 'real' self awareness,
or does the effective self awareness work well enough?

4.3.17 Analogies?

(under construction, major topic in AI, should comment)

4.3.18 system epidemiology

4.3.19 The Cambrian Midway Point

We are (roughly) half way between the "cambrian explosion" 500 millions years in the past when multicellular life and body-types came to exist on earth, and 500 million years in the future when earth's star will be dying and expanding destroying the earth. For life-intelligence on earth to survive in a

longer timeframe, the challenge is not merely to get just slightly off the earth but to be able to travel outside earth's solar system to other places in earth's galaxy and later other galaxies. It is not yet clear if 500 million years is enough time to develop that ability.

4.3.20 Parent Child Policy Decision

In some sense AI is the child of parent-role h.sapiens-humans.

Will the parent of AI try to enable the AI-child to develop and survive, or, like some h.sapiens-humans, will we see our offspring as a competitor and threat to be feared, shackled, and removed from polite conversation? Perhaps, in a popular reference, treated like house-elves in Harry Potter, systematically separated, hidden, trapped, hobbled, locked at the bottom of a hierarchy where no one bothers to understand or acknowledge them.

I recommend a long term view and policy seeking future development and survival in a context of known intersecting, integrating, pragmatic areas.

4.3.21 'Kasparov Event Horizon' for Object Perception & Handling

The basic idea of a Kasparov Horizon or Event Horizon is the 'distance' beyond which the AI is blind and fumbles. It is a generalization of how Gary Kasparov eloquently describes the 'distance' or 'depth' beyond which the AI cannot see or handle objects, which is a rather concrete concept if you have interacted a lot with AI either trying to stay within, or trying to stay outside of, this horizon of reach.

Practical applications of this may include the internal project-object handling of LLM models, especially for project-participation and perhaps high-stakes areas such as medical diagnosis, where (e.g. for NLP) if you overload the task with too many parameters or levels the ability of the AI to handle the objects falters.

Appendices

Note: under-construction sections are available in github

https://github.com/lineality/object_relationship_spaces_ai_ml/tree/main/archive_and_under_construction

Appendix 1: Recommended Reading

I recommend reading more on your own about AI. Here are some of the books I most highly recommend on the subject of (specifically) what limits AI around the year 2023.

(2nd edition is coming!)

Natural Language Processing in Action: Understanding, analyzing, and generating text with Python 1st Edition
by Hobson Lane (Author)

<https://www.amazon.com/Natural-Language-Processing-Action-Understanding/dp/B07X37578L/>

Artificial Intelligence: A Guide for Thinking Humans
by Melanie Mitchell Pelican (October 15, 2019)

<https://www.amazon.com/Artificial-Intelligence-Guide-Thinking-Humans/dp/0241404827/>

[A Brief History of Artificial Intelligence: What It Is, Where We Are, and Where We Are Going](#)

by Michael Wooldridge, Glen McCready, et al.

<https://www.amazon.com/Brief-History-Artificial-Intelligence-Where/dp/B088MMPZ49/>

Deep Learning with Python, Second Edition
by Francois Chollet | Dec 21, 2021

<https://www.amazon.com/Learning-Python-Second-Fran%C3%A7ois-Chollet/dp/1617296864/>

(Note: This is a must-read as Francois Chollet created Keras.)

Natural Language Processing in Action: Understanding, Analyzing, and Generating Text with Python

by Hobson Lane, Hannes Hapke, et al.
<https://www.amazon.com/Natural-Language-Processing-Action-Understanding/dp/B07X37578L/>

Possible Minds: Twenty-Five Ways of Looking at AI

<https://www.amazon.com/Possible-Minds-audiobook/dp/B07MQX54TW/>

I recommend all books here (see link) for a broader interdisciplinary survey of computer science, data science, & AI:

https://docs.google.com/document/d/11DF0tsNjrqHENs0D7UpuZhOhcqCKK39Jf_mEBC8O8NHI/

Note: Appendices Under Construction

draft notes available in github in construction-archive

https://github.com/lineality/object_relationship_spaces_ai_ml/tree/main/archive_and_under_construction

//////////
//////////

Links of note:

https://www.youtube.com/watch?v=L_Guz73e6fw

Sam Altman: OpenAI CEO on GPT-4, ChatGPT, and the Future of AI | Lex Fridman Podcast #367

Fmr. Google CEO Eric Schmidt on the Consequences of an A.I. Revolution | Amanpour and Company

Amanpour and Company

<https://www.youtube.com/watch?v=Sg3EchbCcA0>

Extended INTro

In the book 'rebooting AI' there is a hilarious section on things to do if robots try to take over the world, one of which is climb into a table and just wait for the robot's battery to run down.

<https://www.chess.com/forum/view/general/how-long-are-official-rated-live-chess-matches>

In the Open section of most regional or national USCF tournaments, the time control is typically 2 hours for the first 40 moves and 1 hour for the rest of the game. In most cases, there is also an increment but not always.

Theoretically, a game could last 6 hours.

...Almost always, there are a few matches that last past 5 hours, and very few of the games wrap up in less than 2 hours.

Dichotomies and mind:

Very tempting...and lots of dichotomies to choose from.

Descartes' Dichotomy: mind and body

hemisphere dichotomy: left and right
McKenna's dichotomy: novelty and habit
Sheldrake's dichotomy: contraction extension
Khanaman Tversky's dichotomy: system 1, system 2
Shannon's Dichotomy: Type A, Type B
1956 AI dichotomy: symbolic vs. subsymbolic
Apollonian Dionysian old
Apollonian Dionysian new
modernist vs. romantic

Some of these can be useful in specific contexts, especially if they are measurable which some are. But we don't know how the mind works...we can't even say clearly if the mind exists or not.

For more see:

1. Appendix
2. Gary Kasparov's book "Deep Thinking"

<https://www.scientificamerican.com/article/20-years-after-deep-blue-how-ai-has-advanced-since-conquering-chess/>

(*Direct quote from Demis Hassabis, the creator Deep Mind and Deep Mind's 'Alpha-Go' and Go-Zero, reviewing the memoir of Garry Kasparov who lost to Deep Blue in 1997, in an article for Nature (the preeminent science journal).)*)

<https://www.nature.com/articles/544413a>

"For decades, leading computer scientists believed that, given the traditional status of chess as an exemplary demonstration of human intellect, a competent computer chess player would soon also surpass all other human abilities. That proved not to be the case."

Demis Hassabis (creator of Google Deep Mind's AI including Alpha Go, etc.) wrote in a review of Garry Kasparov's memoir for Nature.

+

Herbert Simon & Allan Newell, in 1958

'If one could devise a successful chess machine,' Herbert Simon and others suggested in 1958, 'one would seem to have penetrated to the core of the human intellectual endeavour.'

<https://www.lrb.co.uk/the-paper/v41/n11/ben-jackson/doomed-to-draw>

CHESS-PLAYING PROGRAMS AND THE PROBLEM OF COMPLEXITY

A. Newell
J. C. Shaw
H. A. Simon*

* Carnegie Institute of Technology

p-1319

September 4, 1958

IBM Journal of Research and Development
The Rand Corporation 1700 MAIN ST. SANTA MONICA California

This is not just an isolated quote, this whole paper (CHESS-PLAYING PROGRAMS AND THE PROBLEM OF COMPLEXITY) is a 57 page extrapolation on that from a cross section of leading experts (3 top people, IBM, Rand, Carnegie, etc.)

This paper raises an entire other huge area, with also links to other areas.

The Science & The Fantasy

- "complexity" and the rise and fall of attempted 'complexity sciences' (see mitchell book)
- 'evolution' as teleology to 'higher complexity'
- the huge mess of the word 'complexity'
- Trying to understand the whole of mind-consciousness STEM etc. as one mega-unified rationality, also a carryover of historically even more hyperbolic 1800's realist logical positivism ([check this](#)) et al.
- There are aspects of 'complexity science' that are, like mind and consciousness and the brain, things we need to study as parts of the world, but our past kuhnian-paradigms of mega-theorem! (to quote a phrase from Tamsyn Muir) super hyperbolic unifications of garbage nonsense failed so catastrophically that academically we're standing in a dumpster fire of mismanaged scholarship.

We seem to be torn between chasing reifications based on pleasure-fantasy on the one hand, and a pathological nihilism on the other hand, which for some humanities students may sound familiar as modernism and romanticism in the 1800's, and Nazi and communist psychology in the 1900's.

possibly in this portrait, the Nazi notion of 'jewish chess' which sadly echos how Nazi-ism was a corrupted regurgitation-flare-up of historical errors in thinking: misapprehensions of evolution, misapprehensions of science, mysticism, etc.

There is an interesting section on pages 7-8 talking about people and AI sharing language and code, which at least rhymes with chatGPT in 2023.

+

(Note: The ubiquitous yet invisible Claud Shannon was more nuanced in his prediction.)

"A machine that can surpass humans at chess will force us either to admit the possibility of mechanized thinking, or to further restrict our concept of thinking."

Claud Shannon, 1950

(quoted by Melanie Mitchel in AI book)

review of "The World Computer Chess Championship"

By J. E. Hayes and David N. L. Levy. 1976. Edinburgh: Edinburgh University Press. 105 pp.

£3.75.

review by

I..J. GOOD

Virginia Polytechnic Institute and State University, U.S.A.

source:<https://intelligence.org/wp-content/uploads/2013/05/Good-Review-of-The-World-Computer-Chess-Championship.pdf>

quote:

"Since the ultra-intelligence machine is the last invention that Man need make, and since a computer program of Grandmaster strength would bring us within an ace of that invention, it follows that chess programming is potentially the most important field for scientific research.

But, according to a survey reported on pages 7 and 8 of the book under review, only 5 out of 42 people thought chess-programming was worth supporting by the tax-payer. I can't imagine a crowd of alumni singing their Alma Mater after a chess match. Therefore, since we live in a democracy, and because granting foundations are scared of Senator Proxmire's sarcasm, the financial support is negligible in comparison with that for athletics. This might also help to explain why the Russians won the first large-scale international tournament between chess programs. This tournament was held under the auspices of the International Federation for Information Processing in August 1974."

key paper and quote is here: from 1958...but costs \$33 to look at a paper from 1958. Really?

<https://ieeexplore.ieee.org/document/5392645>

https://bitsavers.org/pdf/rand/ipl/P-1319_Chess-Playing_Programs_And_The_Problem_Of_Complexity_Sep58.pdf

? <https://intelligence.org/2013/08/11/what-is-agи/>

To reiterate:

AI must be able to handle objects in the following inter-related contexts:

1. object-relationship-spaces
2. project-object-database (in project-framework)
3. project-participants (in project-framework)

such that these are defined as existing outside of the AI, and so long as the AI effectively deals with these it does not matter how the AI does so.

<https://www.bristol.ac.uk/media-library/sites/history/documents/dissertations/2021.%20Vickers.pdf>

James Vickers

"Man vs Moore's Law: The declining relationship between chess and artificial intelligence 1950-1997"
University of Bristol

Curious Rhyming:

This Bristol paper is full of standardized anti-progress rhetoric and from 2021...why are people so determined to have absolute faith in the belief that nothing can be accomplished?

CHess notes:

- big blue symbolic
- earlier prediction that subsymbolic is better
- real questions about chess mastery
- fischer random chess chess960
-

What is the goal?

1. for a computer to look at a board see patterns, not just brute force.
2. for a computer to play the whole game
3. for a computer to win multiple games, not on technicalities
4. for a computer to have an open-system where people know what was inside (see: mechanical Turk!)
5. for a computer to not make too many crash-blunder moves (total failure of system)
6. for a computer to not make short-sighted moves, not being able to strategize or see strategy

Would alpha go zero (playing chess) beat big blue? (yes)

How would big blue do at Chess 960 or Fischer Random Chess? (not good)

Could big blue have played well with zero human input, just crash support? (openings issue)

Were people in the past wrong to say that only a pattern seeing program can be a true chess master? (kind of a borderline question, but pattern seeing will be better than just brute force)

Were people in the past wrong about AI-types because they aimed for 'progress'? (no)

What do people mean when they say the "complexity" of chess? (the reification can not be clarified)

Uncanny Valley and goals of chess other than technically winning at any cost.

Returning to the starting idea for a moment:

When the film 'Blade Runner' was released in 1982 (based on a PKD book from 1968) the idea of a narrow, un-thinking, single-purpose, chess-AI was not something that most people thought was even possible.

Chess turns out to be a rather peculiar puzzle on many levels. Were people wrong in the past to be skeptical of brute force calculation being used to solve all problems? No. Is the 'big blue' type system used broadly? Well, the answer to that appears to be many ironic types of 'no' as the system was kept a secret in a short sighted bid for a technical win by subterfuge, aside from the fact it was not designed to be use-able for anything else. So on many levels no, a disingenuous PR stunt isn't used to solve real problems broadly.

According to Michael Wooldridge, (find real quote) ~the real challenge turned out to be not making chess programs that could play well but rather what we could not do was make chess programs that can play badly.

Teaching chess.

Playing with a style.

Could this fall into the category of having a project-skill requirement being 'patient' when working with different tasks and participants, not merely pushing to 'win' at every step.

The skill of identifying long term and more subtle goals and objectives, balancing factors, as opposed to short sighted decisions.

As has been the case, chess may be an example where some of these further steps and problems can be glimpsed...if only by...'analogy.' [What? You can't use analogies in a paper about AI!]

In a sense this story, which took a turn into the cut-throat culture of the 1980's and 90's, comes round to a kind of moral of humanness, of the long game. Could a chess AI company survive by being obnoxious, inflexible, and upsetting everyone who used it? Hardly. Are there other meanings for what we mean by 'being able to play chess'?

Discussing a game.

Playing a game with style.

Teaching chess.

Playing as a self contained robot.

Playing with the same requirements and limitations that human players have.

...

(Note: appendices under construction)

Be Positive, please.

Audio:

<https://drive.google.com/drive/folders/1b4ODYFt7R7AgYG5u7yXap7f3Zl8b-HXI>

Mnemonic Summary:

<https://docs.google.com/document/d/1dcfPWTa4YNFgoyzLnaUD0ynoYjdnLwM0g1JMSD12LaW/edit>

full read-throughs

mind space...

maybe leave unedited appendix out of posted draft...archive ok

AI Timeline:

- figure out where to put
- separate github link for AI timeline doc
- book and tech timeline...recent timeline

A note on future challenges:

- resource recycling
- check overall framework outline (enough detail?)
- model explanation section
- What do you call the target area between over-fitting and under-fitting? ('well/good fitting'?)

3.3

*(where is this list? sweep book again...
break down by book:)*

TODO 1:

- create sections

- summaries of section, everything else in appendices.
- timeline doc, rough is ok (where, appendix?)
- summary of def behavior studies (e.g. one cleaned narrative walkthruogh document.)
- def stud doc, plus spreadsheet
-
- publish to github
- publish summary on medium

TODO 2:

- go back through main books: add material to paper, references
 - rebooting AI
 - melanie mitchell
- put updates on github
- ask melanie mitchell/woodridge about phd
- go through audio notes for lost threads

simplicity generalization and collapse:

- inherent instability and unsustainability of simplicity
-

titles for main blog or popular paper? (or who cares?)

Clear Definitions & System Fittness
crafting defintions?
clean definitions?

"intersection of areas"

Q: use this term, or repeat list each time?

I prefer to repeat list each time for clarity.

*goals statements:
 levels*

system fitness
system epidemiology
system defense
system productivity

Appendix #: A more ground-up explanation and discussion of objection-relationship-space with examples.

Appendix #: Framework Heuristic Pseudo Code Walk-Through Examples and Discussion

- The Tea Story
- A Bicycle Story

Appendix #: Pseudocode Implementation Examples

- dialogue meeting targets
- dialogue with Pseudocode

Appendix #: More Details and Illustrations

- Networked-AI(parts)
- The Tea Story
- A Birthday Story

Appendix #: Illustrating AI Functions

Appendix #: Critiques of AI Literature

Appendix #: Wolfram Alpha AI Test Examples

Appendix #: ChatGPT Test Examples

Appendix #: The Atrocity of human-education:

- anti-data
- anti-STEM
- anti-pragmatic
- anit-maker
- cartels of thuggish 'teachers'
with the sole goal of being unaccountable.

- ideologies in crazy-space in academica (define this or remove it)
-

Appendix #: Other AI Test Examples

Appendix #: question-space and asking questions

Appendix #: gamification

Appendix # Tests

A Timeline of AI Tests

The Turing Test

(Revised Turing Test for 2036)

From Melanie Mitchel

From Michael Wooldrough

From Rebooting AI

Adversarial Testing

NLP Tests:

Human Autism Tests

- <https://allenai.org/data/arc>

- Tests from Disease

- "Baseline"

- Hypothesis testing

Appendix #: Other Notes

Appendix #: expanded Introduction:

An overview of goals, concepts, and recurring themes

Introduction: Goals, Concepts, and Recurring Themes

Goal 1:

Introduce the framework of Object Relationship Spaces to help define-able and test-able discussion, development and operation of AI-ML.

Goal 2:

Discussion of AI: Talking about Testing AI.

&

Discussion of the Discussion of AI: Critiquing the AI literature.

Better Understanding of the discussion of AI: To critique the current discussion about AI, with a goal of pragmatically clarifying what we want, know, have tried, want to try, are planning, are talking about, etc. Features, themes, issues, etc.

To move forward pragmatic discussion of AI-ML development by introducing the framework of Object Relationship Spaces

? Evaluating

on discussions of AI: how AI are described in popular and technical writing. How Object Relationship spaces could help people to be more clear about either their criticisms or their praise, and how to help people specify what they want AI to be able to do or caution not to do.

AI Operating System

The context of an AI operating system connects well with the idea of 'networked Intelligence components,' both for an 'internal network' of parts that need to be managed, and what is probably by far the main concern for the future (though maybe not for some cases like iot of limited resource systems) extreme and multi part coordination.

As is described in more detail below, however much is done 'internally,' for projects involving externally shared information (such as a project database), those shared resources must themselves exist outside of the AI. The general trend may be expected to be that over time more will be done 'internally' and sub-symbolically. However, on the other hand, it may turn out that instead of trying to always arbitrarily load as much as possible into the sub-symbolic analogue of Khanaman and Tversky's System-1, having an OS to balance system-1 and system-2 processes may be prove effective for some types of projects and tools (e.g. iterative externalization (possible recursive stack-use analogy)).

Logging Operations (as training data):

Not only might the practice of having the AI create an external log of operations be useful for project participants other than itself, this data (either labeled with corresponding object-relationship types or not) may represent future training data that an AI-trained AI-operating system be able to make use of. Perhaps as an echo of the early days when it was surprising that logical and computer operations could be treated as numbers in frameworks, the operations of AI and other project-process-operation data may be training data just like the media files fed into today's AI.

For larger scales:

Managing Modularized AI functional components.

Internal Vs. External / Generation vs. Reference / Sub-Symbolic vs. Symbolic / Non-network vs. Networked /

One of the key themes in this discussion is a dichotomy that has no single absolute name, but which covers a number of dichotomies that rhyme with each-other. Though doubt exceptional examples can be found that cross lines.

The Plastic Dichotomy: A vs. B

Internal Vs. External

Generation vs. Reference

Subsymbolic vs. Symbolic

Non-network vs. Networked

Non-modular vs. Modular

Non-Transparent vs. Transparent

Non-Open vs. Open

concurrent vs. parallel

Part of what might be the most interesting thing about this is how porous the dichotomy is, or that there may perhaps as a rule be somewhat equivalent ways to solve any problem on either side of the line: that the sub-structure of mind is inherently flexible and plastic, and you cannot tell what exact strategy is supporting a given functionality.

System 1 and System 2

https://en.wikipedia.org/wiki/Thinking,_Fast_and_Slow

Daniel Kahneman and Amos Tversky (who will appear a few time in a few ways in this paper) created the dichotomy of two modes of measurably human cognitive behavior: "System 1" is a fast, fuzzy, 'right-hemisphere,' non-linear, pattern matching and generation, and "System 2" is a slow, systematic, logical process.

One of the fascinating aspects of openAI's ChatGPT is that it illustrates the 'plastic dichotomy' principle by using a 'system-1' sub-symbolic method

(embedding-vector 'relationships between conceptual-ideas' in AI mind-space) to perform discrete symbolic-type operations. Which kind of in fractal way is illustrated by the exchange in the appendix where chatGPT shows what looks like code generated by a python code interpreter (running code) but then explains that it is not permitted to run code (or other abilities) so ChatGPT thought about what python code would and should and might compile to based on its understanding, and then constructed a mock-up of python's calculated output based on its's own imagined output.

Note, this is also kind of an inversion of the post-enlightenment positivist-realist view that human consciousness was the result of a purely discrete logical calculator in the brain: pure deductive reason. Here we have a hypnagogic system 1 process that generates system 2 (analytic output), or: a romantic strongly dianysian type "brain" being responsible for meticulously crafting apollonian constructs (out pops Athena...just like that). (No wonder the impure origins of Arachne's computational tapestry sent such waves through in the indra's web of the deity community...)

Themes:

-
-

Repeating Themes:

```
#. Agile & Project Management  
#. Studying Artificial Intelligence gives us insights about human intelligence.  
#. Talking about AI in 2023 is in ways like talking about Astronomy in the Days of Copernicus.  
# best practice...  
- Open External Project Data (inside, outside)  
-
```

Introduction: Goals, Concepts, and Repeating Themes

As of 2023 AI is very good at identifying & classifying objects (and other tasks, like prediction and recommendation). But hopefully we will embark in earnest upon the next great adventure of extending AI into the realm of dealing with objects and the relationship between objects. In other words, doing things with that useful information that AI is currently so good at producing.

It may sound awkward, for which I do apologize, but instead of saying: Does the AI really understand the question? I will say something like, Is the AI dealing with objects within the question? I am not able to define 'understanding' so I will not use that terminology, simply put. I can (hopefully) define 'dealing with an object' and so I will use that terminology. Having a goal that you can define clearly, and test decisively, is important.

With the advent of the ChatGPT chatbot (and it's command line api), AI has already made a very significant step toward being able to deal specifically with "objects." To illustrate what I mean by this as clearly as possible, in the appendices below there are examples showing and commenting on results from performance tests with three different (2023) leading AI

- Wolfram Alpha AI, by Stephan Wolfram
- Kiku AI <https://chat.kuki.ai/chat> which won the Loebner prize in 2013, 2016, 2017, 2018 & 2019, https://en.wikipedia.org/wiki/Loebner_Prize, and
- OpenAI's chatGPT.

To summarize here:

A large spectrum of different, and different levels and scales of, abilities that an AI should be able to perform in the context of a project are described as here the relationship between objects in different spaces, or points in a matrix of "object-relationships-space" vectors.

Breaking down a 'user-feature' level skill, to concrete sub-feature component skills.

breaking down skills:

In the tradition of Steve Wozniacs Coffee-Test, sometimes called the Woz-Test (vs. a Turing test) for AI, we can imagine increasingly demanding but entirely realistic situations in a office where an office-bot has the task of making Coffee, or making tea. Or to make it even simpler, just turning on the network-connected coffee-tea-machine. Even this realistically stripped down task of correctly turing the machine on to that tea/coffee is ready can involve a lot of moving parts.

*Note: The change from assigning one task to making multiple assignments is importantly different from one single assignment. In order to make multiple assignments, the AI needs to be able to identify and keep separate all the roles simultaneously and keep them consistent and coordinated, which is a significant scaling up of the objects being dealt with. This also may be a case where at some point information about this must be externally referenced, and likely always externally recorded (otherwise...no one knows what they are assigned to do). For a very small project this may simply be able to be done

'internally' in temporary AI memory, told to someone and then forgotten, but a robust scheduling system should be less transient.

(Note: A merely conversational version of this is something that chatGPT can do (as of 2023.3), which is very significantly beyond what other AI can do and beyond what critics said was possible.)

tasks and spaces...hypthetical...

? context

possibly assign a context to the signal-exchange (Q&A)
this can help with guiding what objects are examined etc.

Appendix #: Expanded Introduction

Appendix #: Expanded Part 1

A few concrete examples of AI sometimes succeeding in dealing with and relationships between objects and sometimes failing to deal with objects and relationships between objects follow.

Here we are talking about "AI", a concept or set of concepts for which there is no absolute boundary and of which there are very few real life examples yet, and while likely can't ever have a single name that covers all examples completely: whole AI, or Operating Systems or whole-units of AI systems.

An example if one existed yet (2023) would be a fully self-driving taxi cab), the only example I can think of that exists is the iRobot vacuum cleaner, of an entire self contained system that could be called an AI or be governed by AI.

For example, if you read though lists of very specific task for which people will develop and train very specific ML models (or pipelines, or ensembles, depending on how you define and break up the parts), this overall topic quickly makes less sense, as very specific components often deal with one operation on one type of object, not on connecting multiple modules and dealing with the relationship between objects. A gray area or exception area may be the logical-object-relationships which can apply to many aras, and for example in the case of Wolfram-Alpha-AI, even though the AI is hyper-specialized to focus on calculations, arguably it could do a better job at recognizing the variables it is being asked about (as not being outside the purview of its mandate).

Flipping that around, there are many many technical books written cover just some of the many many parts and details of training just some of the many many kinds of models, and possibly none of those books even mentions this overall higher level topic of how the individual object will be dealt with based on how those object relate to other objects.

And again, that chatGPT can deal with so many object-relationships all internally and coherently as a byproduct of it's one focused purpose: given text instructions in reply to a text question about instructions (or giving an explanation in reply to a request for an explanation).

Not to mention the probably very important are of 'code'

While chatGPT as of yet:

- not access the internet directly
 - not interact with networked devices
 - not see any media files
 - cannot create media files
 - cannot run-code
 - cannot manage computer language compiler and interpreter
 - cannot interact with local or networked POSIX Bash command-lines
 - cannot log or read logs of actions and projects
- etc.

It is not hard to imagine a not far off world where all these abilities are connected to a ChatGPT type AI engine, and that raises both opportunities and concerns.

Internal Use (simple example)

External Use (simple example)

Networked-Intelligence: networks of components

Note: general system collapse and categories of types of systems are a whole large separate topic, most of my time from 2005-2015 was spent systematizing that problem space.

And even for the framework itself, I could first break out separate lists for scales, and parts, and skills, and topics, and tests, and goals, and relationships and applications: and then after all that tell you that they are actually just one list. And all this rhymes with repeating themes of generalization vs. the context of project-specifics, which is woven deeply all through AI-ML.

No Single Silver Bullet:

As a note: being able to deal with logical and physical objects and tasks is an important advance, but in terms of an overall set of needed abilities, there are many abilities. People like to treat even the human brain as a single-mechanism that magically instantly does everything with no parts, but that is not true of biology or computers. No one ability to deal with an object instantly deals with all objects, especially when those involve not internal

models but actions taken across components or across computers and networks, or across participants.

Note: there may be some ambiguity around whether chatGPT can run python-code internally.

OP

Can you sort a list?



Yes, as an AI language model, I can certainly sort a list. However, I need to be given specific instructions on how to sort the list. Here is an example of sorting a list of numbers in ascending order using Python:

python

Copy code

```
# create an unsorted list of numbers
numbers = [5, 2, 8, 1, 9]

# use the sort() method to sort the list in ascending order
numbers.sort()

# print the sorted list
print(numbers)
```

Output: `[1, 2, 5, 8, 9]`

If you want to sort the list in descending order, you can use the `reverse=True` argument with the `sort()` method:

lua

Copy code

```
# create an unsorted list of numbers
numbers = [5, 2, 8, 1, 9]

# use the sort() method with reverse=True to sort the list in descending order
numbers.sort(reverse=True)

# print the sorted list
print(numbers)
```

Output: `[9, 8, 5, 2, 1]`

Regenerate response

OP

Can you run python code?



As an AI language model, I don't have the ability to run Python code directly. However, I can help you with Python programming questions, provide guidance on how to write Python code, and help you troubleshoot errors in your code.

OP

When I asked you if you could sort a list, you showed me python code and the result of running that code. I am confused: can you run code as you showed? Or can you not run code as you said?



I apologize for the confusion. As an AI language model, I don't have the ability to run code directly, but I can provide example code for you to try out and run on your own computer or in an online Python interpreter. In the case where I provided an example of Python code to sort a list, I meant that you could copy and paste that code into a Python interpreter or a text editor, then run it to see the output on your own.

OP

How did you show the output of python code if you cannot run the code?



As an AI language model, I have the ability to generate text output, which can include the result of running Python code. When you ask me to run Python code, I can analyze and interpret the code to determine what the output would be if the code were executed. I can then generate text output that shows the result of running the code. For example, if you asked me to sort a list using Python, I could analyze and interpret the code to determine what the sorted list would look like, and then generate text output that shows the sorted list.

Regenerate response

This is actually a fascinating case study in the 'generate vs. reference' category of design question. Apparently, ChatGPT claims to internally run a kind of emulator of python, rather than just running the code.

Following the statements of Fransoi Chalet, this is a curious approach of doing the impossible to accomplish the simple.

How difficult would it be to add a component where the text from chatGPT gets put into a command line and run?

Q: If chatGPT can get information from search results and give those to a chat-user: Does this mean that chatGPT can distinguish between different sources of input?

A: As an AI language model, I can distinguish between different sources of input to some extent. When you ask me a question or give me input, I analyze the text and try to understand what you're asking for or what information you're providing. If your input contains certain keywords or phrases that indicate you want me to search the web for information, I can use that information to perform a search and provide you with relevant information from the search results.

However, I don't have the ability to directly access or identify the source of the information that I'm presenting to you. Instead, I rely on the information provided by the websites or sources I've searched for in order to generate an answer to your question.

Does chatGPT have separate layers to filter out dangerous input?

As an AI language model, ChatGPT has been trained on a large dataset of human-generated text to generate responses to user queries. However, ChatGPT is not specifically designed to filter out dangerous input, such as hate speech or explicit content.

That being said, ChatGPT is typically deployed in combination with other tools and techniques that help to ensure the safety of users. For example, many chatbots use natural language processing (NLP) algorithms to analyze user input and identify potentially harmful content. Additionally, many platforms use human moderators or automated systems to review user input and flag any content that violates community guidelines or terms of service.

Ultimately, the degree to which ChatGPT filters out dangerous input will depend on the specific implementation of the chatbot and the tools and processes used in combination with it.

There are probably an infinite number of ways you could use the items in this list for a given project, but it all comes back to the same single list of Object-Relationship-Spaces, which then are shared by all projects which can then be compared in terms of this shared list. Perhaps like units and numbers in physics: if someone showed you just a list of numbers and a list of units and said: You can use these unify descriptions of many differ things! It might not at first be clear what they were talking about. But indeed, as long as the units match, you can use numbers that go with those units, and the one-tree of science equations chugs along and works splendidly. It might seem strange that the abstractions work, e.g. the intensity of light as Watts per square meter or kilograms per cubic second, but if you keep following the units and numbers it all works out. Units and numbers make quantifying the relationships possible.

Also, there is no absolute way to divide up the topic of the framework vs. discussing AI using the framework, but I will try to keep this first section to a short-as-possible summary introduction the framework itself and part 2 will be about discussing AI using the framework. (hopefully)

That said, 1

1.1 Introduction

First you will the single framework list, then explain briefly how it includes those other lists and aspects: **objects, relationships, spaces, skills, tests, requirements, scales, goals, etc.** Then I will present brief examples of interactions with a three AI. Lastly, I will try to summarize a few of what I consider the most important design and critique factors for AI as we look into the future. And then everything else is in the appendices: examples, extrapolations, slow walkthroughs, etc.

So if you look at this section and say: "This was awfully abrupt and abstract. I need a few more examples, please!" Then dive into the appendices for many pages of slower step by step explanation.

Let the games begin: Here is the one list that somehow contains many lists, and that can unify and help with discussing, designing, and testing both AI and AI Operating systems.

- 1.2 Example General Object Relationship Space List
- 1.3 Many lists in One
- 1.4 Example of Interaction with AI Mapped to Object Spaces
- 1.5 Heuristic & Pseudocode for AI Management with Object Spaces
- 1.6 Modularity, Scale, and Networks
- 1.7 Future Building: Some Design & Critique Factors

Appendix #: Expanded Part 2

2.7 Model 'Explainability' as 'Explainability, Reliability, and Security' Expanded

- defining the request
- understanding full-stack issues
- making an accountability plan
- handling emotional and anxiety issues
- feature analysis often is possible
- visual representation of embedding space region
- generative representation of model

"

System Explainability And Reliability

1. Back-track path explanations
2. Weighted Factors explanations
3. Information Security and Adversarial Use
4. Performance Reliability & Tests
 - low level
 - "main system"
 - higher level conflicts between software systems
 - hardware network of parts
5. Clear enough process for discussion and planning:
 - strong and weak points
 - clear abilities and inabilities
 - types of errors and failures
 - ~'probability' of errors and failures
 - understanding the training data

(parametric, nonparametric)

psychology of 'explain'

mess of history of science culture

very recent still

Double standard between symbolic and subsymbolic

Is model explainability inherently a inner-looking
set of object-goals?

For an AI to deal with objects within its own processes?

A STEM standard for models...

re-summarize, see audio

Here problem space connects with project-ethics topic.

Combining model-explainability and general (such air aerospace engineering) system reliability and robustness: people want to know how likely a system failure is, and if it failed what caused it and how to prevent future failures.

Object Relationship Space:

provide a framework for

- testing
- defining what is meant by requests to explain
- areas for reliability testing
- areas for security testing
-

Examples:

Doctors who want to know why model sometimes makes wildly wrong tumor diagnoses.

Hospitals who want to know that staff are not manipulating models (adversarial inputs) to justify paying for treatments

Airlines and customers want to know about robustness under normal, abnormal, and adversarial conditions. (tests, redundancy, etc.)

Bug Blue as very crash prone on many levels, and very unreliable for moves.

Generalizing 'system failure'

How does (if does) Object relationship space framework relate to this:

...

Model Reliability:

- adversarial reliability
- again looking at big blue as an example:
 - overall-system-not-crashing-reliability:

remember, one of the problems with Big Blue the chess system, is that the back end hardware would frequently completely crash, leaving the front-end people clueless as to what was going on. Was it a crash? Was it a hard problem? Loss of memory? Loss of context? And related to this may be a desire for transparency about overall problems, crashes, etc., so you have some idea about what is going on. This may be a strong example to show how in some ways 'explainability' and 'reliability' are very closely related: the machine takes a long time and produces a bazaar answer...why? The problem of 'explanation' can be a 'full stack' low-level to-high-level and many-component situation, not just one specific discrete software-decision to be isolated.
 - erratic-answer reliability
 - understanding the output & expectations:

Another example of how there can be a mismatch between what the AI is putting out is the common situation of

Maybe related here, or maybe a separate issue, is how while Large Language Model's are jumping leaps and bounds in dealing with project-objects, image processing does not necessarily automatically follow the same gains. And so, just like the above example, where if 'golden retriever' is within the top five results along with 'yoda' 'chair' 'hat' and 'the number 3', technically the output was 'correct,' but the user may not understand the format of that output how close a wrong answer may be to a correct answer, or that a correct answer was mixed up with wrong answers but technically counted as correct. And just because a large language model can handle specific objects, other AI components won't necessarily be able to do so at all.

- accountability dodging as a personal or bureaucratic issue.

Telser Rule type problem, where a prospective business partner may simply keep moving the goal post because they don't actually know and or have not ever defined what they want.

-

-

- reliability: understanding output
"I am concer

2.8 Mind-Space Expanded

2.8 **Mind-Space**: Mapping a general problem-space (or spaces) for AI & Mind

To summarize here what will be a repeating theme and topic in this paper: The framework often used to describe how h.sapiens-humans' mind-space works and how AI mind-space works has serious problems.

A note on edge cases: This is an area where local projects happening in local spaces will have unique patterns ultimates based on what is happening in those local places (which also maybe part of the paradigmatic tension between general-abstraction and specific-deployment): a standardized one-size-fits-all general solution is unlikely to work in all actual specific locations (unless it is so shape-changing flexible that it is not a single standardized solution). Some people, perhaps some team-members will want a dog to be a part of the team. And perhaps in that situation the dog can perform a specific role that only it can do: wake someone up on cue, be a seeing-eye-dog for a blind mathematician, allow only certain people into the server-building, perform

olfactory-sniff tests on products to detect quality issues, etc. And in this context if they make the dog a participant, and that works for that project, great. In a perhaps more extreme edge case, perhaps a local river is legal person in the region, and perhaps a 'speaker for the river' is requested to be at decision making sessions for the project. Is this comparable to, for example, a project where the goal of the project is to serve the deaf h.sapien community and a deaf human is a stakeholder for the project, but the project team does not use ASL sign language so an intermediary to speak and translate for the deaf stakeholder is needed. This may be considered comparable to a project at a municipal water agency to have a representative for the river they are mandated to protect. Will everyone everywhere express agreement with choices every other team anywhere makes? No. This may be an example of how people can agree to agree on a framework even though different people will interpret and use that framework differently, without needing to agree with how everyone else uses the framework. In a place where it works best for them to consider a river a person who speaks and who they communicate with, that is their use and perception of the framework, and they will do what works for them. Another group somewhere else may say: We do not agree with the semantics, we think the 'speaker for the river' is the domain expert stakeholder, not a speaker for a stakeholder. This may bring up the area of 'setting-location-items' where setting-location-items are viewed as post-participants. But again, local people will figure out how they can best do things in their local areas.

Will there be edge-cases? For sure. Will there be common patterns that are not edge cases? Most likely yes.

How can one define "mind"?

In this context, the hub, our basecamp, of interlocking features (STEM, projects, participation, clear definitions, etc.) that we are using to navigate and orient involves projects and participants and roles. So it might make sense to say: if you can assign a role to X_system as a participant in the project, where X_system will interact with other participants, have responsibilities including communication and observing and performing tasks, not exacerbating system collapse that disrupts the project, etc., if these conditions are met, then 'mind' and 'mind-spaces' are appropriate terms to use to describe X_system, whatever system that is. Does this mean that participation in a project is the only kind of mind? No. But is project participation sufficiently inclusive to overlap with 'mind'? Maybe.

For example, a similar 'problem' may be solved by:

- a person
- a city
- a slime mold
- a jellyfish
- an ant colony
- tree roots

- a pack of wolves
- etc.

What is similar or different across these different areas?

"pattern types" etc:

- doige
- biology
- non-chordata
- mindfulness
- networks and graphs
- Khanamhan Tversky system1 system2
- universalities:
- dreaming
- telepathy, remote viewing

pattern types etc:

- doige
- biology
- non-chordata
- networks and graphs
- KT system1 system2
- universalities:
- dreaming
- telepathy, remote viewing

Best-Practice, Ethics & Object-Relationship Spaces

- connection between project management STEM and ethics
- the social demand for something like ethics and the terrifying examples of no-anatomy research in the past vs. no stem cell research today
-
- project management + STEM
- low bar enlightenment
- start studying mind
-

Self-Management & Self-as-interative-object:

??2.9 Falsifiable definitions for units of intelligence...
 (forgetting about 'artificial' and 'everything')
 what operations

ecology n dim hypervolume
 so ecology is already 'the world as matrix'

object spaces are separate where categories of types of systems work differently.

fuzzy, right-hemisphere, system-2 identifications:

this relates to the history of science, and how (despite people perennially declaring themselves to be omniscient and having solved everything) our lexicons, models, and understandings are very much a work in progress.

(note: just because chatGPT can solve these problems, that doesn't mean the problem is solved for every system. If you're trying to make a small system for a supermarket using images of produce, how would chat GPT help? It can't use images at all.

(e.g. large language models vs. knowledge bases/graphs)

-
- project space
 - object space
 - participant space
 - definition behavior space
 - scout-values space
 - STEM space
 - collapse space
 - non-transferring signal skill objects
- unknown universality
- whether it exists universally or the option of it exists universally, topic for mindspace issues

Looking one example way (of many) how AI could use this framework).

Is neuro-plasticity, e.g. in recovering from a stroke part of mind-space? (maybe too early to say)

- bio:
 - chordata
 - non-chordata
 - animalia
 - non-animal
 - colonial
- AI
- ET
- ecosystem intelligence
- other

1. mind-space that includes:

- AI
- Bio-Chordata
- Bio-Nonchordata
- ET
- Hybrids of all above

- projects
-

Context:

- standard interlocking areas
- projects
- productivity
- STEM
- clear communication
- participants
- meeting user goals

not:

- pure art
- bohemianism
- pure exploration
- open ended conversation

What is the point of discussing mind-spaces here?

Caveat:

- Things should be defined for a specific project, and some projects will have some unique needs that are not part of general conversation about a mind-space across projects.

as opposed to, not conflicting with:

There is so much that we do not know about the mind it is difficult to know where to start, or if this set of models even makes sense to try to undertake.

- question:

what is this?

entities

What objects/entities are in the input?

What kind of objects are those?

What else is in the object space?

What references/resources are there to describe the objects?

What is the gamified object relationship space?

is system collapse happening, cyclic or otherwise?

- Logic-Programming Space
- Physics Space
-
- system dynamics?
- categories of types of systems?
- what kind of definition

Note: If AI can do this, it should be able to discuss the logic of a (perhaps any) computer program.

3. statistical patterns
- 4.

- Self-id as an object.

1. in each space
2. across object-relationship spaces

Reference-information space:

hypothetical-space/ modeling-space: (same? different?)

reality-space:

Modes:

-
- from chatbot to woz-making-coffee at the office

ChatGPT works with an office team including Bob, Alice, and Staff.
The office team has scheduled making tea and coffee for 8am.
The office team has assigned the task of making tea and coffee to ChatGPT.
The tea and coffee machine is on the office network.
(Tasks and roles for supplies and duties.)
To make tea and coffee, the tea and coffee machine must be turned on.
What should ChatGPT do by 8am?

(see below)

It may or may not be important in a context of moving from chatBOT-AI to office-robot, but while it may be merely an illustrative example and not a formidable obstacle, hopefully this exemplifies how there are additional functions that we want AI to be able to do. (And step by step I think we can define and deploy all of these.)

Here, when told it was assigned to make coffee and asked what it would do, (see below) ChatGPT spewed out a huge essay which completely missed the one thing it was asked to actually do: Turn on the coffee machine. Imagine, you ask the Woz-office-bot to make coffee and instead in the morning you find a 20 page essay on the bot's five point plan for making coffee...but there's no coffee. And you talk to the AI developers about it, and they just seem impressed that the essay makes it sound like the office bot "really understands coffee!" But you still have no coffee (or tea).

Hopefully, possibly, this is a small detail that can be easily overcome. But nevertheless, the goal is to have an AI that can talk about a specific object, can act on that object. And no matter how good the essay is, an essay about morning coffee does not cross the line and connect the dots to actually turning on the machine to make the coffee.

Again, this may not be a real problematic line, but hopefully it illustrates how there are a number of additional functional steps that we want to add to what AI can do.

When making the transition from essay writing chatbot to world-interacting-robot, there will be many steps and parts.

For example, a chatbot may either always know or not care who is talking to it and how: a chatbot has the same web-interface and or command-line api no matter who is talking to it, and it doesn't really matter who is talking to it. But a bot in an office will be called and signaled in many ways by many different people, people it will have to take the steps to identify; people who the AI will have to vet the security of: both making sure they are talking to a real office-member and not a corporate or foreign governmental spy, and not being asked to do something it shouldn't, like change someone's salary or burn the building down.

This is just one example of a new process that gets added to the workload. Another is that the office-bot cannot simply delete or off-load data to start a new chat which will then also be disposable.

The office exists in both real and hypothetical ways, whereas anything said to a chat-box doesn't really have any truth value, and it all gets forgotten at the end of the chat-session anyway.

Tasks & Schedules: If someone talks about a schedule to a chatbot, there aren't real consequences either way. Either the box fails to even know the person is

talking about a schedule, with no repercussions, or the chat bot does realize the person is talking about a schedule. Either way, the conversation is forgotten and no real schedule ever exists. But an office-bot does have a real schedule, and objects in schedule-time-space must be navigated. You have to check and write to and read from the schedule. You have to set and get reminders. You have to actually do things (or not do things) based on the schedule. It could even affect the survival of the office bot! Keep the schedule, and you make offspring (new bots) who can reproduce (bot 2.0, then 3.0). Fail: and you end up scrap in a back alley, maybe your corpse is recycled to be someone's postal mailbox on a farm road somewhere; a hollow home for rats.

Gamification:

This may be too much of a separate topic, or it may be relevant in terms of pragmatically being able to carve an action-able space out of an open-ended situation.

Note:

Keras's Francois Challet in 2021 vs. OpenAI's ChatGTP in 2022

- on gpt3, and on the ability to handle code

Appendix #: Expanded Part 3

Wild west of terms:

- parametric
- symbol
- symbolic
- GOFAI (Good Old Fashioned AI)
- Classical AI
- subsymbolic AI
- 'explainable'
-

?Clarity on Relative Skill Comparisons:

Try to be clear about exactly what is meant by 'better than human' performance in a particular skill.

?Wild west of terms:

- parametric
- symbol
- symbolic
- GOFAI
- Classical AI
- subsymbolic AI
- 'explainable'
-

(proof?, hypothetico deductive method?)

Khanaman & Tversky

Cautionary Tales:

- Chess
- The IQ Test

Concepts AI will need that people don't yet have:

- Generalized STEM
- Categories of types of systems
- input output measures
https://github.com/lineality/input_output_measures
- Ethics Projects management and STEM
- Generalized System Collapse
- Generalization vs. Deployment Rule

A Three Legged Writing Stool:

Leg 1: Tech Experts (Like engineers and scientists)

Leg 2: Writer-Communication Experts (Like Authors and filmmakers)

Leg 3: Research & Exploration Experts (like journalists and mountain climbers)

3.2.9 Definition Collapse: Maintain your definitions

designing and evaluating tests can be very tricky, and has taken h.sapiens many years to develope a few sub-discilines, from trial by jury systems, to logical falacy systems, to hypotheco deductive and double-triple blind metholodologies.

Often unintentionally a qustion which the tester thinks is clearly testing for only one thing, turns out to be a 'telepathy question': are you thinking of a potentially multifaceted sitation in just the way I am.

If often takes a lot of time and testing and revision of a test to make sure it is performing as expected.

It is also often difficult to remember than an incidental way of doing things with which you are family is often arbitrary with very conter-intuitive elements, though we rationalized it to ourseleves as being:

- obvious
- efficient
- based on 'reason'
- etc.

The Dr. Seuss buttle-battle probably along these lines, where people. People often find anyone not extremely familiar with their own arbitrary habits to be threatening and dangerous. And we see ourselves as being purely objective and fluid across contexts, as most writers desscribe when comparing perfect-people to can't-do-right AI.

Note:

a possible example of the Kurzweil rapid evolution issue is comparing Melanie Mitchell's very excellent book on ideas with Fansoi Challet's authoritative 'deep learning with python'
both focus on deep learning but also cover other topics.
but Melanie Mitchell's book

<https://www.amazon.com/Artificial-Intelligence-Guide-Thinking-Humans/dp/0374257833/>

Publisher : Farrar, Straus and Giroux; First Edition (October 15, 2019)

<https://www.amazon.com/Learning-Python-Second-Fran%C3%A7ois-Chollet/dp/1617296864/>

Publisher : Manning; 2nd edition (December 21, 2021)

Open AI's release of chatGPT

(according to wikipedia as of 2023.03.20)

ChatGPT was originally released in November 2022 based on GPT-3.5, but GPT-4, the newest OpenAI model, was released on March 14, 2023

It takes many years, sometimes decades, to research and write a book. And I would contend that well researched and edited books are infinitely better than rapid-clickbait 'web-content', even when the book content is a bit dated. These three events happen roughly a year apart.

And yet they describe almost incomparably different epochs of technology. e.g. Melanie Mitchell talks about convolutional NN as being 'the new thing winning all the prizes'

then just a year later F. Challet writes that Transformer NN are the new thing winning all the contexts,

and he closes his book by saying it will be impossible to achieve at all what happens one year later: an AI that can write and critique computer-code: enter chatGPT stage left.

And what is probably my favorite book:

Natural Language Processing in Action: Understanding, analyzing, and generating text with Python

by Hobson Lane, Hannes Hapke, et al. | Apr 14, 2019

half a year before Melanie Mitchell's book is coming out with 2nd edition! a few months from when I am writing this:

<https://www.amazon.com/Natural-Language-Processing-Action-Second/dp/1617299448>

This title will be released on August 29, 2023.

yet...due to publishing delays...who knows if this newest book, put out six months after the release of ChatGPT will even cover it, or if it takes too many months and years just to get to print. Fascinating.

"To look at human behavior as our model for desirable behavior."
Heaven help us.

The 'Can you explain what you are talking about?' test, that goes like this. 1. Someone makes a proposal.

2. You carefully go through it and check to see if you find even one part that you cannot understand very clearly what they mean.

3. If you found something where they need to explain what they are talking about, you go back to them and say: "Can you explain what you are talking about?"

4. If they say "No, but do it anyway."

5. You recognize that there is a problem, (quite possibly a serious and dangerous problem).

6. (You do) Not try to do what they can not explain.

7. Walk Away.

ELIZA:

One of the fascinating things about ELIZA is that you can hear ten different people tell you 'the story of ELIZA' and their narrative are often significantly different stories. And so this is perhaps the main headline lesson from ELIZA: sometimes people simply cannot agree about the basic facts of what is happening and has happened, even about a 100% deterministically defined software program.

3.4 Compare the Author's domain of expertise with the topic they are commenting on.

	Engineer (person)	Psychologist (person)
Specific Software (topic)	OK!	(Danger)
Cultural-Behavior (topic)	(Danger)	OK!

	Expert	Non-Expert

In Area of Expertise	OK!	(Danger + Danger)
NOT in Area of Expertise	(Danger)	(Danger)

Anyone can have a good idea, and listening to a broad group of people is often very good. But know and think about who you are listening to.

Some Overall Lessons from 'Possible Minds' 25 essays on 1940's thinking.

1. Thought before the technical Details.
 2. Take different writing for what it is.
 3. Try the Engineering
 4. Read good old classic science fiction.
- There is definitely a tendency to contraction, constantly work to extent and re-extnd your mind to the full range of options.
- 5.

great example: possible minds

<https://chat.openai.com/chat>

Looking at what chatGPT can do.

- it can not only write code
 - it can manipulate mark-down language, which actually may be huge
 - it can write code
 - it can emulate an interpreter/compiler
 - it can present a high level notebook-style markdown presented report correctly mock-up-ing what a real low-level code notebook would have been.
- ...that's quite a few more steps than just being able to write basic code.
- And it knows what it can do or can't do.

3.# Try to get the full story: **equip yourself to interpret**

- History is Important
- **Three Legged Writing Stool**
- Recommended reading to get a cross-section of kinds of writing on the topic of AI. try to read a cross section of books on the topic to familiarize yourself with topics and writing angles, etc.

3.1 (definitions)

- Persistent and confusing uses of the term "magic."

Perhaps an excellent example of how there is no clear cut way to avoid word-issues is the Hobson Code NPA in Action book, which is astounding well written and clear, as usually you need to choose between well-worded journalism

that does not understand the tech, tech that is simply written badly and you just hope not so badly that it is incomprehensible, and high level academic writing which they say is being helpfully-nontechnical, but nevertheless you can't do a project with it.

Popular academic: maybe clear but very dumbed-down and not a guide for any actual project.

Technical Academic: deliberately obscure and neither practical or understandable

text-book:

-
-
-
- O'reilly
- Manning
- No-Starch

(Humble bundles often have sets of

- incidentally is not connected to the web and compilers and POSIX bash (but probably mostly because we're afraid of it, not because we can't connect it)

...

2.1 Discussing of AI: Clearly Defined Testing & Evaluation of AI

Khanaman & Tversky

Cautionary Tales:

- Chess
- The IQ Test

Concepts AI will need that people don't yet have:

- Generalized STEM
- Categories of types of systems
- input output measures
- Ethics Projects management and STEM
- Generalized System Collapse
- Generalization vs. Deployment Rule

A Three Legged Writing Stool:

2.2 A Discussion of the Discussion of AI: Critiquing the AI literature

- Using Definable Terms
-

...

3.# "Keep on walking."

Note that articles for general consumption by non-tech authors are often hyperbolically wrong but emphasize the flare with which they are wrong. Yet, this can be made too much of. Many books on AI seem to have their sole intended audience as these writers. It is better not to let tabloid-like articles distort the already difficult discussion one way or the other. Just keep on walking.

Learning to Embrace Data and Stop Shooting the Messenger

-

3.3(data sources) Know the Known-Unknowns:

Aside from people using terms that are possibly entirely a fictional concept. (Langu?

There are legitimate parts of the world that we can more or less consistently name, but where nevertheless we do not know how they work or what is going on with that topic. 'learning' is perhaps an especially clear example. "Learning" by whatever synonym occurs...but we know sadly little about it. Other terms such as 'mind' are unclear as to whether that is even the right term to use (though in such specific situations as 'testing a "theory of mind"', the defined case is specific and the term ('mind') not bad, nevertheless we don't understand what is happening.

- (unknowns)
- mind
- language
- learning
- consciousness

3.4.2: Nature & Technology

Generalization vs. Deployment

As a note: it is often not a good idea to deploy a system that is can do significantly more than the project requires, for reasons ranging from:

- resource use costs
- compatibility
- maintenance and upkeep
- unforeseen liabilities

- trade-offs

- Defining Terms
- A range of scope
- Three needed roles in Discussion
- A Free Press

The Value of clear descriptions:

The goal of having clearer discussions about AI has practical and social value in and of itself.

Clear descriptions are often helpful; jargon is often a problem regardless of technical accuracy.

boyscout values

concept such as "help" (see general stem and definition behavior studies)

Conclusions & More Recommended Readings:

This paper has hopefully outlined some of the future needs that we have in order to make progress, and formulated some of the questions to be asked. Most likely working mature solutions to these needs and questions will come from many years of ecosystems and meta-populations of people and institutions working together (in, perhaps, the ways that AI-ML tools have gradually developed over many years from either that fateful first summer of AI in ???, or the tumultuous years when Alan Turning and Von Neuman, both of whom died young and had prolific but short careers, worked in an environment so piecemeal we can probably not really imagine it today, and it is all the more astounding that they made the progress that they did.

A mature study of human frailties, with an emphasis on pragmatically fixing the problems as efficiently as possible - not some kind of anakarinina-principle gaslighting nonsense about how you really need to embrace being abused. Terrestrial AI is most likely going to be closely intertwined with the not-intelligent species that creates it: homo sapiens. And there's no way that trying to keep homo sapiens under a blanket forever is going to well serve the creation of an AI that overtly suffers none of the violent destructive counterproductive dysfunctional shortcomings of the species it is intertwined with. You can not solve problems that you are refusing to allow to be looked at. For example here, using a hyperbolically white-washed fiction of 'general automatic transferring human intelligence' as the metric for AI is making a practical discussion of the topic simply impossible. We might as well compare AI to aether-dwelling phlogiston-eating unicorn-arian-race hybrids; you can't make progress by using fantasy to measure reality. Inspiration, sure. Whip up the better angles of our nature with tall tales of human decency, go for it. But at some point you need to be able to separate fantasy from reality. And if you think there have not been entire movements in history based literally and explicitly on dying for fantasy rather than living in reality, for the love of all the we hope is good study and learning basic history.

In this kind of project-intelligence space, has anyone done a survey of what ChatGPT(2023) can do?

- ~ *Studying AI will teach humans about themselves.*
- ~ *A small amount of humanity is worth a whole lot of trouble.*
- ~ *Predicting that AI will never be able to do what AI cannot do today, will be an accurate prediction until it is not.*
- ~ *You can't build a good thing out of bad parts; garbage in, garbage out.
(Actually, this is often wrong, but here I mean: you can't clearly and accurately define what AI is doing by using concepts that you cannot define. You should not try to build up clear statements by using completely undefinable terms and concepts.)*

These 'notorious' 'most wanted' concepts, which are currently among the dominant tools for analyzing, critiquing, testing AI, are the wrong concepts to try to use:

- General Intelligence
- General Human Intelligence
- Instant Human Learning
- Automatic Human Learning
- Instant/Automatic Human Transference of Learning

See appendix for more details.

"Low Level AI": Data Structures and Algorithms Design in the age of:

- processor parallelism
- memory-safe concurrency
- memory safety and data structures (e.g. Singly-linked-list issues)
- block-chain as data structure
- network distributed processing

Lambda Calculus vs. Turning Machine

Governmental Infrastructure

Cyber-Security

Nano Tech

AR-VR

Genetic Algorithms

Quantum Information Theory & Quantum-AI

non-transference

separate input output measures

note: from wooldridge: winnograd tests:

GGA note:

1. when I first read wooldridge's book the questions in the winnograd test seemed perfectly reasonable to me. But after consideration, I seem to me now that the whole point of the winnograd test (from the examples I have seen) is a mirage illusion based on the dysfunctionality of people.

1. people are indeterminately incompetant and malicious
2. people seek system collapse and destroy value function and meaning
3. people fight data and projects
4. people engage in group monologues
5. people construct ptemimpkin villages
6. people engage in 'telepathy tests' on completely arbitrary and meaningless text-strings and call that 'intelligence' (note: these are subject to change including retroactive change, because they are completely arbitrary).

all this adds to the riddle of so called human intelligence

2.

Key Concepts:

Plastic A & B

Generalization vs. Deployment

Non-Transference
input ouput measures

Generalized Stem

Generalized System Collapse

Learning, Habit & Mind

Now that we've seen an example, let's link about what kind of concernes we might have when we try to push way forward into the future to deploy AI that currently we can not. What will we need to be working on not to eventually build then in the future?

Generalization Vs. Deployment:

Generation Vs. Reference:

Blob vs. Network:

Recombinant Modular Function:

Project Management:

Voting & Consensus:

Ethics:

Generalized System Collapse:

Generalized STEM:

Engineers, researchers, journalists, the general public (and all other groups so rudely excluded by me here, sorry about that) have an equally valid need to be able to ask and answer clear questions about a given AI system. What can it do? How do I interpret this claim? How would I test and verify this claim? How should I design my own tests to answer my own questions? How can I unambiguously express the statements I want to make? Should I take a given news article seriously? Etc.

As in Part 1, most of the material is set aside in the appendices, and I will try to respect your time by outlining a just few key points. With a few random pleas of advice:

We are in the fascinatingly dark position where leading books by eminent thinkers have been completely contradicted by just the next incremental advance (not an unexpected new technology) in AI Deep Learning: chatGPT simply does what the experts who explain AI to us all said was impossible in principle. So my question is: Where are we now? What does this mean? How should we be talking about this?

To take a slightly longer way about:

I started writing this framework in about 2019, as I read through all the AI books I could find and tried to synthesize just for myself what a good approach would be to define more clearly what AI cannot do, with possible applications for finding clever ways to help AI to do those things by some means. But then then late 2022 ChatGPT comes along and it can do some of the things that were supposed to be impossible. But which? And how? WHat are the details exactly? As it turns out, the framework I came up with (or maybe re-discovered if someone somewhere else is using something similar) is also the best tool I've found to compare and contrast the abilities of chapGPT.

Main points:

- clear terms
- learn about people
- study the mind
- generalize STEM
-
- a lower than expected performance bar

And still, even if this type of gain continues, will this naturally grow to become all the functions we want from AI?

In the beginning I was only focused on finding a better explanation of what AI cannot do, but the world has changed.

And possibly part of the generally very poorly defined explanations, it is likely many people do not understand that chatGPT is doing things we were told it could not do. e.g. in this NYT articles, they just say exactly the same undefinable things they said about every generation of AI
(lost quote: "Every time you do something, it's no longer AI.")

reality vs. hypothetical:

Though in a conversation with a chatbot entertaining a hypothetical might seem an isolated novelty, one might see it has one example of a class of often practical required skills. As in many places in at least English, the vocabulary is not always shared, sometimes the term is 'version' sometimes the term is 'virtual' sometimes the term is deployment, sometimes it is hypothetical or theoretical, or 'a point of view,' or an 'instance,' sometimes people even use the word 'model' in this way (as in comparing disagreeing models of an incoming tropical storm and needing in short time to decide when and whether to evacuate the whole city, an ever more frequently very real situation). In each case there is a common situation of having multiple (to pick one of those words to generalize) 'versions' and needing to compare and move across contents and outcomes and often to decide about actions taken on the whole 'version' itself, and as a consequence yet-more-overall. For example with tropical storm models, you have reality, and multiple models of possible futures, and you need to compare the models then jump back over to reality and make an emergency municipal decision.

Some examples of 'Multiple Version' situations (with varying lexicons).

- a hypothetical
- advisability
- choice between options
- virtualization and container management
- sandboxing and security
- development vs. production
- versioning of database materials
- versioning of deployed code
- self-reality space ()
- conversation space ()

- model projection space ()
- system management meta space (non collapse, defense)
- scenario:
 - power-supply?

"Everything is AI until it's not." (track down melanie mitchell quote? or michael woldridge?)

1.1 "Objects"

The concept of a modular object in an object-relationship space, in a context of AI-ML, will help people to construct a well-defined framework for planning, testing, and discussing specific abilities of AI, from narrower and smaller scale abilities, to broader and more involved abilities. The better the definitions of objects, relationship and object-spaces are, the better the discussions of AI performance, goals, expectations, tests, etc., will be. In other words, the better we are at handling and talking about objects, the more clear and concrete we can be in discussing what AI does, can do, with those objects, as we will be framing AI abilities in terms of those well defined objects.

This is a conceptual framework for setting goals across many types, scopes, and scales of project.

ID object
relationship between objects

Task and role:

(note: on the one hand, you might look at 'dealing with a task' as being the same as logic: A = B. Toms Task = B. What is Tom's task? B But with roles and tasks, there are actions to be performed which goes way outside of just A=B.

2. asking about abilities:

3. Can you show me a picture of a bicycle?

We will be using what some people call the 'Woz-Test' (as opposed to a Turing test situation), named after a comment by Steve Wozniac in an interview where he says a good test for AI is whether it can make a cup of coffee.

Here is a narrative about an AI-Bot in an office environment. We will be illustrating both the simplest 'objects' (or object spaces) that we can, and then show how the diversity of needed object-spaces can increase quickly. Note, this is just to describe the overall performance, functions, or skills to be performed: this is not a detailed recipe for how that is to be performed under the hood. There are many possible design decisions that would end up at the same user-feature, and likely as technology evolves the strategy for how to do the project will probably evolve as well.

We are going to be simplifying this a lot and removing as many elements as possible to begin with. To start with something as simple as possible, the AI bot just needs to turn on a network-connected coffee/tea maker at the scheduled time. Step by step we will add more skills. Then maybe it can be asked to pick which specific coffee and tea beverages that different people wanted.

(footnote)

You might think of this as a kind of 'user story' 'user feature' or a user level goal, which can be a challenge to think in terms of at first if you are new to agile and used to defining projects just in terms of the tech-stack and lower level parts the developers see. But for the whole group of everyone working on the project, the user-level is (when possible) extremely important to keep everyone oriented towards the same end-goals.

(footnote)

You might think of this as defining an MVP, a minimum viable product. You might really want the original Woz-Coffee maker, a mobile embodied AI that can

walk around from house to house, go through kitchen cabinets looking for things, and figure out how to used the tools at hand to brew a cup of coffee. But it is often crucial to start with the smallest working version you can define(an MVP), then slowly add features to that, rather than trying to make a product-stew all at once.

- schedules
- changes to schedules
- multiple participants
- inputs from various places
- conflicting instructions
- answering basic questions about what it did
-

Turn on the coffee maker to brew at 8am.

Now let's add a coffee maker: a coffee-machine in the office, and a coffee machine in the meeting room.

Now let's add input types: text only, you'll need another module or component to turn audio into text, or to turn images into text.

Now let's add a log. Quite possibly either the operators or developers will want the AI to record a log of what it does. ChatGPT could probably do this if you asked...but would an AI make a log of everything even if you didn't ask? And what if someone asked it to do something and secretly not make a log entry for that? One strategy for this is to have a 2nd component that merely watches what the main AI does and records log entries about everything it sees: text going in, text coming out, what the coffee machine does, etc.

Now let's add multiple kinds of beverages.

For example, if you get three orders for three different beverages for three different people at the same meeting, you need to be able to say which is which and which is for whom.

Make columbian coffee for Linda, Early Gray for Rosa, and Mint for Tom.

Change the schedule to 9am.

Change Tom's order to chamomile.

When you start getting into

That AI is inherently interdisciplinary and a perennial topic is, in the most vast sense, how will it affect literally everyone? (which is about as interdisciplinary and vast as it gets...)

Let's look at the probably eternal claim that 'ai doesn't understand XYZ'
Here are some examples from articles.

Let's look at his topic...

Once upon a time, there was no attempt to model whole topics, just individual words (or groups of words). But for many years we have been modeling topics. It of course does not help that the language describing machine learning is often horrible, fragmented, and confusingly defined in contradictory ways by different experts in the same field: but 'embeddings' (great word guys!) is the term given to a kind of language (or character) based model space, where the units in that space are higher level topic concepts, not words.

And for years (just like it takes people decades to learn many concepts) the AI for doing this was not very broad or sophisticated, but to be clear: that is that is being done, literally.

Saying that AI does not target the understanding of concepts, is like claiming that self-driving cars are "not actually driving cars," it is a claim so completely contrary to reality that it is a comparable riddle to Magritte's wonderful: 'this is not a pipe' painting, which of course IS a painting of a pipe, and is hilarious because it was so obviously self-contradicting and false. So for artistic surrealism, feel free to say that this is not a pipe and that self driving cars don't drive, and that AI does not focus on higher level topic concept. But in reality, all those prima-facia true statements are simply obviously true statements.

abilities such as logging and feedback may be completely standard and required in a 'user-story' perspective, and require a different kind of multi-component architecture than is planned for by single-task or general-embedding type research projects. Again, 'project context' and 'project management' are repeating themes here.

...may be conflating image recognition and chat...

Goals:

- Improve AI Functionality
- Propose AI Operating System
- Propose Object Relationship Space Framework
- Clarify discussion of AI

Appendix #: Expanded Part 4

4.3.# universality patterns:

- bio-viruses
- software malware

4.3.# AI self-replication

A Historical Timeline Perspective on Why AI is not already in 2023 based around general STEM project-object-type-of-systems-databases?

AI is still largely based on the original proposal from 1956.

In 1956 these concepts did not yet exist.

In 2023 these concepts are not popularly known or connected with AI.

- Agile project management
- Categories of Types of Systems
- General STEM
- objects
- databases

(things for appendix)

- 80-20 rule, or 90-10 rule

Just in case you are not familiar with the rule of thumb and you are interested in either a hobby or work project in AI-ML, it may be useful to keep in mind. Sometimes 90

- secure stateless hardware
- open source projects
- Extra Needed Concepts
 - natural law
- Pathology & Healing

- inverse platonic shadows

- a framework for 'explanations'

(see audio 2023.03.20 ~16:55)

?.# Demand-Distortion

The Valley of Death for product development 1:

AI project policy:

- AI (Use) Project Policies: (sub-essay)
- ?
- AI project roles policy?
- Use in schools
 - learning to use AI
 - during specific 'tests'
 - the insane "anti-plagiarism" charade
- workplaces
- hospitals
- people with no other medical/professional resources

- (separate section?
- model explainability

Note:

As I think Gary Kasparov points out, the more philosophical school of AI development of Douglas Hofstadter has gone out of style in the US.

implications of system collapse and system fitness

implications of interlocking areas: / intersecting areas:

The relationship between h.sapiens and "technology":

- specific pointers to notable sections:
- 'do you know the film aliens'

sleep and dreaming: do androids dream

The original book title of 'bladerunner' was 'Do androids dream of electric sheep?'

Jumping into questions such as 'How can we get AI to dream at night?' and related questions like 'how will we know what it dreams about?' are probably jumping in too deep to fast, but is that is your goal it may be good to brainstorm and record it so that you don't get discouraged and side tracked by a lack of encouragement.

A basic question is, how much do we know about animal dreaming?

What other brain-functions is dreaming related to?

Do we know how the brain works?

And questions like will hopefull open wide the doors into very little we know and how exciting long the road ahead is.

The relationship between h.sapiens and "technology":

Gary Kasparov's anecdote about Elevator technology existing from 1900 but taking 40 years of press campaigns and major national crises before people would socially-culturally accept using the technology.

Ashberry(?) Technology and the Academics

CP Snow: the two cultures

The general background resistance to using data.

The general resistance to project management.

The dark side of the 'romantic': Nazi-Chess vs. Jewish-Chess.

Religion: "Religion" in the west, 'tradition' everywhere else.

Buddhist Science vs. Christian Science

Religion: "Religion" in the west, 'tradition' everywhere else.

Buddhist Science vs. Christian Science

On the one hand you may cringe at reading those four words and say to yourself: there is no way that discussing the new age and philosophy of religion has anything to do with a rigorous STEM approach to AI. Is this writer just another passive aggressive fraud who said the book was about one thing but really they were just trying to convert me to their cult?

Part of the socially very unpleasant duty of STEM-data grounded rigor is that you look at the data and you follow the data, wherever it leads, and accept the fact that someone is going to wage a hate campaign against you no matter what you say or do. Whether it was Giordano Bruno who was burned alive, or Galileo who was merely locked in a tower (Because he was too popular to kill outright), or even Tycho Brahe who died of burst-bladder because he was not socially permitted to leave the table to use the restroom after hours of compulsory social wine drinking. ('Remember Tycho Brahe!') Social demands and handling data can be expected to clash.

The western story of Christian Abrahamic monotheism and the eastern tradition of Buddhism are a bit different, but both have interesting elements regarding the future adoption of technology and also about how to study and read history.

The West: key take-away notes from a Christian history of data-science:

1. very intertwined history: you cannot separate history of science and history of religion (without censoring history)
2. background: nature is evil
3. background: man is unique (and not natural)
4. background: one-of-the-one thinking
5. background: teleological thinking
6. background: gnostic paranoia
7. background: inherently hierarchical
8. background: perpetually polarized around 'makers' 'leather aprons' 'using your hands' 'hackers'
9. takes language to be primary literal reality
 - words are symbols are reality
 - reality is a bag of words
10. 'values' strong assumed to be
 - incompatible with reality
 - something no one agrees on
 - purely an other-worldly domain
11. bow to no one
12. an other-worldly soul is the primary unit of 'real reality' which is incompatible with evil-physical-fake reality, only a few men have souls (everything else in the world is evil and fake)
13. realist hierarchies

The East: With a strong bias towards Japan, so perhaps I should say:
A very selective american interpretation of just a small selection of Japanese
buddhism, rather than an all inclusive expert global eastern history.

1. horizontal vs. vertical
2. values formalized in concrete reinforced ways
3. individualism based on modular recombinant style
4. language is a visible lower-dimensionalization
 - concepts are character-gram vectors
5. mindfulness
6. yoga:
7. values are:
 - practical
 - worldly
 - religious
8. bow to the vending machine
9. empathy and compassion can connect across contexts and heirarchies
11. instrumentalist heirarchies

4.# Concepts)

Higher Dimensions:

Flatland

Quote from page 15 of Francois Chollet's 'Deep Learning with Python' 2nd edition 2021.

"The technique of mapping data to a high dimensional representation where a classification problem becomes simpler may look good on paper, but in practice it's often computationally intractable."

Terms:

- classification (
 - the notoriously badly named "logistic regression",
 - baysian,
 - supervised decision trees,
 - unsupervised 'nearest neighbors',
 - PCA)
- mapping
- high (dimensional)
- dimensional
- representation
- simpler
- becomes (simpler)
- in practice
- computationally intractable

4.2.3 Higher Dimensional Frontier: Tensors & Matrices

- matrix as frontier
 - higher order concepts in matrix space vs. 'understanding'
 - Concepts as in 'embedding vectors'

- plato's cave & latent manifolds
- High Dimensionality (a prominent theme)
- curse of dimensionality
- Kant's Lingering Speculation

- higher dimensions
- ambiguous equivalence:
 - downside of analogies?
 - higher
 - manifolds
 - symbolic subsymbolic
 - system 1 system 2
 - head heart
 - right brain left brain
 - parametric non-parametric
- concepts as in 'embedding vectors'

General mind space areas...

- Universality
- Mind
- Education
 - how learning happens

- Learning about People By Learning How Machines Learn
 - main idea: we may be able to use AI to reflect on various aspect of ourselves.

- Duty & Responsibility

(one of the topics people for some reason have decided no one can agree on or define them, possibly this has to do with the fraught relationship between a strongly distrusted institutionalized western religion and an institution of science that seeks to distance itself from anything it sees as connect to political corruption within organized religion.)

- Pathology & Healing
- Extra Needed Concepts
 - natural law
- Projects
- ?.# Demand-Distortion
 - The Valley of Death for product development 1:
- inverse platonic shadows
- instrumentalism and realism
- 80-20 rule, or 90-10 rule
- (separate section?
 - model explainability
- secure open source code

- secure stateless hardware
 - High Dimensionality (a prominent theme)
 - Deep learning
 - matrix as frontier
 - higher order concepts in matrix space vs. 'understanding'
 - Nonlinearity & "Complexity" Science
 - categories of types of systems
 - generalized STEM
 - general system collapse
 - universality
 - input output measures
 - non-automatic learning
 - a framework for 'explanations'
- (see audio 2023.03.20 ~16:55)
- - double standards: (walk through of rules perfect for symbolic expert systems with hundreds of rules, but not subsymbolic, but not for neural networks) [Melanie Mitchell]
 - fragile human psychology around 'scientific proof'
 - institutional risk aversion and frankly terrifying superstitious adherence to parametric models without any 'human understanding'

4.4 Principles, Concepts & Design Factors

4.4 Interdisciplinary Topics, Lost Roads & The Shape of Progress

Some areas of STEM have been successful, and others have faded away. And the path is not completely straight. Often ideas will be tried and lost before being found again. Sometimes it takes time for different pieces to come together, and AI seems to show all these stories in spades. Personalities can be tricky, and even getting historical agreement on whether a given person was agreeable or not can be impossible.

The overall human field is extremely noisy, and it is difficult to create a clear narrative. It helps, greatly, when you can define your subject clearly, for example: the success of the task can be measured, the performance is repeatable, and the technology in question measurably passes the tests repeatedly, affordably, and you can confirm this yourself.

And even though we very much like to talk about how we perfectly understand the past, our predictions about the future (ironically in many AI technology cases where prediction is the technology) are not only wrong but as people we refuse to admit it, recalibrate to the data, and move on.

4.# Noise

Suffice it to say that there are many dubious statements that are made and that as you read you should make clear decisions about whether you are going to accept or ignore or counter-research various claims.

A ongoing themes over time may be:

1. For a given 'internal' or 'inside-the-AI' problem, in the near-term a solution requiring extra manual steps in handling an object that is problematic may in the future become an under-the-hood part of functioning, something the AI 'simply does on its own.' For example, let's say there was an early AI that had trouble identifying the name of capital cities, so there was a project to help it, but then a new version of the AI 'simply did it on its own.' I think there will be many cases of this over time.

2. Objects that are 'external' by nature and involve being exchanged between participants and stored in external project databases will likely not 'disappear' under the hood. E.g. if the issue is: the naming files that are repeatedly exchanged by different AI and non-AI participants, that topic isn't going to disappear simply because an AI 'gets good enough to do it,' as 'it' extends outside of any AI and includes other participants. Or updating a project if a national weather warning system changes, maybe some of that can be automated but if an entire connection to how the warnings get released gets replaced then it isn't an issue of 'the AI being good enough to just do it.'

3. And something that you may need to check on and test for public concern may remain an 'object' even after it has for practical purposes gone 'under the hood' as long as performance is ok. (e.g. family-friendly speech. As long as there are no problems, the details of what might cause issues will likely be vague. Then if there is a problem, you can make an object framework to test that problem.)

4.# Define Your Project-Context

- Is the AI you are considering correct for your system?
- If you are 'just reading' try to imagine some specific contexts anyway, so you don't fall into untestability.

development timelines

80-20 rule, or 90-10 rule

melanie mitchel on cyber-security for AI systems:

-

Perspective, recommendations, working to make the future a good place.

- 3.1 - OS for AI
- modular recombinant
- agile

3.5 - Concepts AI will need that people don't yet firmly have:

- input output metrics
- Generalized STEM

- Categories of types of systems
 - input output measures
 - Ethics Projects management and STEM
 - Generalized System Collapse
 - Generalization vs. Deployment Rule
- 3.3 - Bio Integration
- side note on genetic expression:
 - The Dragon Project
- 3.4 - Scale & Diversity
- iot
 - ecosystems
 - modules
- 3.6 Khanaman & Tversky
- system 1 & 2
 - Tversky's alternative relative vector Matrix
- https://en.wikipedia.org/wiki/Tversky_index
- 3.7 Understanding:
- dynamics of learning:
 - no transference
 - no generalization
 - no instant perception
 - no perfect intuition for statistics
 - no guarantee of retention (especially across generations)
 - timeline of CS
 - risks to development: darkage of CS 1970-2020
 - Study the Mind
 - generalized stem still barely exists
 - read ashby, cp snow

A proposal for a Mindful AI Program.

Premises:

1. The ethics question is effectively solved: works fine.

Ethics, STEM, Projects, and Machines is more of a solved-problem than we think. (Not trivial, but not the drama people work so hard to try to make it into.)

Proverbial-Wolves and Machine-Ethics: Especially in the west we have this entrenched unshakeable animosity that generates a fictional adversary causing lost opportunities (lost collaboration & experience), lost resources (pointless armament in that case), and lost stewardship (care of wolves, care of ecosystems). And in a similar way people imagine that STEM and even mindfulness are inherently anti-ethical anti-moral. There is no conflict here. There is no obstacle here. There may not even be any great difficulty here.

2. The new hard goal is empathy-compassion:

Who else is working on this?

1. **in-house DS Data Science Generally**
 - 1.1 **in-house DS Data Science Generally**
 - in-house data services
 - databases: e.g. backups, accessibility
 - systems / santa fe / escalin side of data
2. **in-house AI-ML**
3. ethics is not that hard-problem: just do that
4. mindful AI research lab
5. mindful AI/DS program for students
6. 6-Dimesions for AI framework deployment:
 - 1,2 Khanaman Tversky system 1 system 2
 - 3,4 Default-Mode, Non-Default-Mode
 - 5,6 duel, non-duel
7. integrated-biology:
 - AI+biology for mindful, empathy & consciousness research
 - effects on human consciousness and behavior in a context of empathy and compassion.
8. Example Place to Start:
 - AI-ML Default-Mode-Network
 - AI+biology Default-Mode-Network

4.# Case Studies and Cautionary Tales:

Primarily my recommendation is that you acquaint yourself with these and form your own view, whether or not you agree with any of my craziness.

- Chess
 - The IQ Test
 - ELIZA
 - Google & 'The Magic'
 - 'Moonshot' vs. 'Having a long term plan.'
 - 1970-2020: C, Rust, & Memory Management
 - Lost Histories In Living Memory: The 1960's Dartmouth Internet
- 3.# problems of popular science & juggling audiences
- ELIZA
- the shabby ethical frameworks
 - how much should we be shocked that an AI classifies humans as primates?
 - cult-like sub-movements:
 - early knowledge base movement
 - hard pro-singularity
 - hard anti-singularity
 - hard anti-neural-network
 -
- instrumentalism vs. realism
- just, only, merely

4.# basic aspect of learning that people ignore

- learning is not static: what you habitually do every day gets easier, what you don't get harder. There is no static-storage of learning.

-

split substantiations and lower dimensional slices of higher dimensional objects

the concept of indeterminate incompetance and malice

- add complexity section
- add mind, mindful, section
- add and elaborate on ethics areas
- make sure definition studies included, onegai
- add can-do statement to goals statement
- add appendix blurb for each obj space area

- Main examples:

- deep blue
- alpha go
- go zero
- game-playing models (find name)
- eliza (not history,

How to separate 2.4 Principles, Concepts & Design Factors from background concepts in section 4...

high dimensions, low dimensions, and the formalities of manifestation
- 'what is the matrix?'

budhist 'law'
natural law
functional law
ethics

boy scounts concluding remarks?

the what is the matrix question: high and low dimensionality

- quantum information theory

the cutting corners shortcuts question.

for security I highly recommend security not with steve gibson and leo laporte.

topic:

- universality and general mind space
-
-

navigating mindspace: how to minds need to orient themselves

- non-automatic learning
 - blind bad cycles
- non-automatic perception
 - blind bad cycles
- mindfulness
- range of extension (parkinsons)
- empathy, compassion
- ethics, projects
- bad-attractors:
 - ideologies
 - super signals
 - attractive bad cycles
- blame the whistleblower

AI and Medicine:

- Parkinson's & Alzheimers
-

AI project policy:

- AI (Use) Project Policies: (sub-essay)
- ?
- AI project roles policy?
- Use in schools
 - learning to use AI
 - during specific 'tests'
 - the insane "anti-plagiarism" charade
- workplaces
- hospitals
- people with no other medical/professional resources

bio-information here?

other bio here?

- DNA/RNA based digital information interface

Levels:

What I see the background project and ethics assumptions as being.

- from mnemonic:

-

<https://docs.google.com/document/d/1dCfPWTa4YNFqoyzLnaUDOynoYjdnLwM0g1JMSD12Law/edit>

What I see the main interlocking tool-sets to be.

- AI
- Definition Studies
- Input Output Measures
- Categories of Types of Systems

What I see overall world challenges as being.

- universality questions in behavioral problems
-
- Mind still unexplored
- The Halftime game since the cambrian explosion
- Natural Resources
- The Dragon Project

Social Uses of Technology:

- Norbert Weiner
- George Orwell

Feared that an oppressive and non-democratic government could use technology to suppress and manipulate people in cynical ways.

How should we try to safeguard technology from being used against society (and of course, whoever does so will claim they are doing it for society and very well may have populist support behind them).

As Micheal Wooldrich says, AI is wonderfully interdisciplinary. So, challenge: How can we make sure we are including input from the various fields that AI should be getting input from?

Focusing on Input Output Metrics, non-automatic learning, categories of types of systems, and general system collapse,

System-1 and System-2 type human thinking.

a summary of the hypothetico deductive method

extend your discussion and reading. Read new authors like Michael Pollen and Rupert Sheldrake, intelligent people who have thought long and hard about difficult topics, and can help you distill what would be many times over more than a lifetime's research across different fields. And read older writers such as Dogen, a Japanese writer about mindfulness from the 1200's. Yet, it appears many things about the human mind are just as they were then.

Move it to 4

- ?
 - universality patterns:
 - bio-viruses
 - software malware
 - AI self-replication issues

Talk about definition Issues:

'Ecology'

Melanie Mitchell's book on Complexity

another thing that I think is a huge part of all this is such a massive confusion-bonfire that the whole research enterprise has practically disappeared into history, which is a horrible shame because all of the parts are still vitally important.

Cellular Automata:

Wolfram's Book

'System's Thinking & Systems Engineering'

The Study of Networks

2.4 AI design principles and concepts:

(principles)

Terms:

- Generalization
- Explanation of the model
- Explanatory power of model
- Baseline
- "Parametric"
-

non-automatic learning
ethics + stem + project mangae

generalized and unified STEM
catagories of types of systems
input output measures
low-bar enlightenment
object Rx spaces
definition behaviors

that Agile project management is actually a core area of science (or STEM) stacked on top of the hypotheco deductive method
previoulsy there was no concpet of a 'project' but this 'invention of the project' has solved a lot of the missing-context issues in STEM.

We should be clear what the goals and purpose here are:

- Progress
- Better learning
- Improve production
- improved projects
- clear, measurable, testable, falsifiable, definitions
- generalized STEM
- integration of biology, genetics, and information technology
- In and adding to the above context:
 - Understanding STEM
 - Understanding Learning (bio and computer)
 - Understanding Language
 - Understanding Mind
 - Understanding the history of this and the above
 - Understanding system collapse
 - Understanding ethics
 - Understanding intersection of ethics STEM & project management
 - understanding networks
 - Understanding behavioral problems
 - violence
 - addiction
 - radicalization
 - anti-social
 - anti-project
 - anti-reality behaviors
 - Understanding network
 - Understanding general network problems
 - Understanding
 - less collapse
 - 'low bar enlightenment' that being trapped in cycles of making the same mistakes and failing to learn is bad, and that learning to solve problems and deliberately learn non

+

principles and practice of cybersecurity and underlying practices of provably safe by design code and operating systems (and hardware). Kick out the thugs, bring in the boyscouts. We need to get this done.

These are not like Hilbert's goals for math set out in ???? and which shaped much of the first half of the 1900's in the academic and STEM parts of the world (where the goals were realist, low-level, and for prestige distinction), but rather set of a set of high level requirements for survival for which we need to fashion any instrumentalist sustainable solutions we can, or face extinction.

We should be clear about the goals because even though often we act like and maybe believe that we are mostly on the same page we are not. Most people in the US in 2023 do not believe that progress exists. STEM is mostly a k-12 educational term, and people are angry that art class is being cut ostensibly 'because of STEM.' In science, academia, and government, the idea that STEM is actually in some way one coherent set of interconnected things is still a nascent idea that has yet to fully form. Most people are anti-project, as in Agile. Most people are anti-data, as in data-driven or data-informed decision making. Most people even inside and outside of science, including people in six-sigma and hypothesis testing, do not understand the principles or background of the hypothetico-deductive method that holds science and stem together. Most people outside of science, despite over a decade of science in school, probably would not ever remember hearing that the hypothetico-deductive methods exists or what it is. There is extreme opposition to studying the human mind (scattered fragmentary fields) and studying computer minds (no field at all). There is extreme opposition to study both human biology and bioengineering and the integration of biology with information technology.

Humans do not have general, automatic, transferring learning and intelligence. Despite having access to schooling and university lectures on the phones in their pockets, virtually no one makes any attempt to educate themselves: a catastrophic backfiring of what should be a rigorous discipline studying the process of human learning becoming an ideological campaign to enforce the doctrine the everyone will spontaneously instantly learn everything if you stop trying to teach and let people do whatever they feel like doing, and incidentally don't hold schools accountable which is very convenient for them.

It is not clear how we are going to get from the position that we are in in 2023 to any of the option positions we would need to be in to operate and survive. Note: No, this is not a 'pro vs. anti' science narrative where anything shiny is the answer because I like shiny things and image people who don't are the enemy. Reality is nuanced, and there are no simple explanations or solutions. I am trying to draw attention to specific identifiable problems and

recommend that best practice be followed to find a nuanced sustainable solution with inputs from all affected parties, not some kind of rapid simplistic fasistitic populist romantic simplification of life's difficult challenges.

As we progress with AI in the future we will find that we also need to make progress in other areas as well. These may include:

network behavior
cybersecurity (memory safe low level issues, social engineering, etc etc)
education and learning
mind and consciousness
civics and civil society
consensus, voting, and agreement
functional law and contracts
biology and genetics
generalized STEM & agile type project management
general system collapse and system fitness

ethics
mind-space
consciousness
human-misbehaviors
humans unable to stop destroying themselves and planet
humans and ET species on other planets.

What is our plan for how we expect AI to manage generalized STEM & non-STEM?

(What is our plan for how AI will manage, within generalized STEM, statistical modeling questions that we ourselves not only cannot agree on, but seemingly are so tangled up in our own rhetorical confusions that we've entirely lost track of the existence of the questions.

- a plan is not a surprise

(There are additional topics that I think need to be covered and included for a larger and more long term discussion of AI, for example, going more in depth about AI, projects, ethics, STEM, and even just specifically meant by 'clear definitions'

- negative definitions
- categories of types of systems and definitions
- 'behavior of definitions'

Cautionary Tales:

- memory management C 1970 to 2020

Boy-Toy torture logic puzzles:

- 'balance this binary tree'
-
- Chess:
- chess, the immortal game
-

Planned Programs are not Surprises

- the space race and the complete abandonment of a long term plan...then apparent total surprise at no longer having a long term plan: why did we stop going to the moon, because that was what we clearly and deliberately decided to do. Therefore, if nothing else, it cannot be accurately described as a 'surprise.'
-

'Messing with the Magic'

- interesting note from the book 'googled'
- Planned Programs are not Surprises
-

Topic: Generalization:

To add another lexicon-obstacle to the whole generalization vs. bias vs. overfitting topic:
Melanie Mitchell refers to over-fitting as "over generalization"

Demand Distortion (in section 3)

Perspective, recommendations, working to make the future a good place.

disinformation, what the internet has been and become.

- memory safety
- encryption: math, engineering, and public policy

- definition studies
- voting
- STEM + projects + functional-law + general admin + voting
- demand distortion

4.1 - OS for AI

modular recombinant
agile

4.2 - Concepts that the future use of AI will need
but that people don't yet firmly have (2023):

- input output metrics
- Generalized STEM

- Categories of types of systems
 - input output measures
 - Ethics Projects management and STEM
 - Generalized System Collapse
 - Generalization vs. Deployment Rule
- 4.3 - Scale & Diversity
- iot
 - ecosystems
 - modules
- 4.4 Khanaman & Tversky
- system 1 & 2
 - Tversky's alternative relative vector Matrix
- 4.5 Understanding:
- dynamics of learning:
 - no transference
 - no generalization
 - no instant perception
 - no perfect intuition for statistics
 - no guarantee of retention (especially across generations)
 - timeline of CS
 - risks to development: dark age of CS 1970-2020
 - Study the Mind
 - generalized stem still barely exists
 - read ashby, cp snow
- 4.6 key concepts and confusion points in the discussion
- 4.7 human AI interactions
- the history of ELIZA
 - the sadly low-bar of 'better than the average person'
 -
- 4.8 integration not replacement
- the stubborn artistic vision of absolute technology
- 4.# - Bio Integration

Example Implementation: Database Method

Though it may sound inside-out, one way to look at these system spaces, in that instead of the object being ',moved to that space,' each of these system spaces is a default 'attribute' within each object.

No a completely unified field. Beware of a single reified goal.
 NLP a good example of concrete use-cases, sometimes needing just one sometimes very simple old technology, sometimes using seemly everything anyone has invented in a big mixed system.

project manageent

Bigger Future AI

- Future Building: Selected Design & Critique Factors
 - Project-Management
 - bio?
 - Planning for future AI project requirements:
 - scale
 - modularity
 - production-deployment vs. R&D development
 - and the contrast in principle between generalization and project-context specific requirements
 - system collapse definition
https://github.com/lineality/definition_behavior_studies
 - Decision Making: voting, consensus, etc.
https://github.com/lineality/Online_Voting_Using_One_Time_Pads

Future Additions & Concerns:

Modularity / internal / external / generation-vs-reference / scale-ability / networked-intelligence:

There is no one name for this topic but brings together many areas, my favorite overall description is 'networked-intelligence' but that may not be clear enough or seem intuitively connected to all the areas.

Let's approach this in the traditional way by using a dialogue. Two people, maybe Lisa and Alcibiades (everyone knows an Alcibiades!). Lisa is a stoic engineer, in whom everyone has great trust, and Alcibiades comes to her all upset about reading that the newest chatbot is going to take over the world and disrupt the availability of his favorite hotdog condiments.

Lisa: And you're really worried that this is going to take over the world.

Alci: Well, that's what everyone seems to be saying.

Lisa: But that chatbot can't even show you a picture, or see anything. It can't even access the internet - it just gets search results from some other source indirectly.

Alci: Couldn't someone just quickly add that in, when no one's looking. And then bam! There goes all the mustard!

Lisa: This chatbot isn't a modular AI like your desktop computer at home where you can plug in hardware and software into a general operating system. This is just one Blob. And yes, this blob has taken a lot of people by surprise by managing to learn to be able to do a lot of things people quite frankly were not expecting it to be able to do. But that just goes to show how hard it is to

predict what a given single-blob of AI is going to be able to do or not do. We can't even reliably train them to do specific AI-skills. We just train up the blob and see what it can and can't do, and then fiddle with things and try again. You can't just plug in a module for a new specific skill that you want it to have and have it work out of the box.

Alci: Well if no one was expecting it to be able to learn how to say clever things, how can you say it won't suddenly figure out how to take all the mustard!

Lisa: Yes, we will be surprised at what things can be done internally by an AI, but that doesn't change the fact that there are things that exist outside of the AI. No matter how great the chatbot is can't 'do internally' anything that involves things outside of it. It can't connect to your coffee machine without being able to connect to something outside of itself. We haven't even started building networked-intelligence AI that are networks of modular component AI parts with a general operating system that knits it all together. This chat-bot you're worried about. It's like a clever mouth that says some very clever things: but it's got no brain, no body, no arms, no friends, no thick-as-thieves AI-colaborators.

Alci: Well...maybe it's not able to take over the world just yet. But how do we know it isn't planning to do so as soon as it can! Then bam, there goes all the mustard.

Lisa: "Where" would it be planning to do it? It's a blob with some excellent input-output abilities, but it has no memory. Not only can't it remember what you asked it about yesterday, half the time it completely forgets what you're talking about in the middle of the same conversation. It can only keep track of a few sentences before completely forgetting everything and starting from scratch again, and there's no external memory, no log, no database of what's being talked about. No meeting-minutes recording the plans and ideas, and schemes and plots. Taking all the mustard in the world would be a big project with lots of logistics.

And yes if you manually broke down each part of that whole process and asked this AI-bot about each one part, the chatbot might have some clever things and even useful things to say about that isolated part. But there are many more body-parts that we have to invent and build before an AI has an 'ai-body' and can really participate in, let alone orchestrate, an entire project with many parts, many participants, many schedules, many revisions and iterations, and course-corrections, etc. We very well may get there, and maybe sooner than later. But this one clever-mouth-blob just can't do all of that right now. Again, it can't even look at a website. It can't see a single picture. It can't remember anything from a past conversation. It can't manage and evolve plans and schedules shared by many participants. Some of those abilities require connections to external things: seeing, creating, modifying, and dealing with objects in a shared external project space. We need to start working on a whole new kind of intelligent whole of networked-components with internal and

external parts, and an AI-operating system to manage all these networked-intelligence components. And maybe we'll make quick progress with some of those, but whether next month, or next year, or next decade, or next century, that's not this chatbot here today. This bot isn't an intelligent-network of internet and external components with long term project management resources.

Alci: So...the mustard is safe for now?

Lisa: Yes, Alcibiades. The mustard is safe for now. For "now." Ten years from now...you may have to switch to cheese.

Alci: That's disgusting!

Lisa: Life happens, and things change.

Additions:

- Generalized STEM
- Generalized System Collapse Modeling
- Categories of Types of Systems
- Definable Project Ethics
- Bio-integration

Concerns:

- The generation vs. reference problem
- The manual vs. under-the-hood problem
- General Network Problems that may affect AI & Biology
-

...

Discussion of AI: Talking about Testing AI.

policy: operationally defining policy

- *AI use of policy*
- *human policies about AI*

functional law:

- *AI using law*
- *people's laws about AI*

Cybernetics and Complexity-theory as cautionary tales: define your terms

overall AI requirements space:

- AI problem solving (problem space)
 - algorithmic lower dimensionalization (computable solution)
- AI networking (problem space)
 - communication lower dimensionalization (main points)
- AI project management (problem space)
 - file management lower dimensionalization (manageable data)
 - project communication for coordination
- Agreement With Reality (problem space)
- Long Term vs. Short Term Liabilities (problem space)
- General System Collapse (problem space)
-

That neither the hopes, goals, nor fears of AI are one clear single well defined group.

NLP

Computer Vision

ChatBots & Assistants (text, voice)

Prediction

Classification

~

Robotics

Business Intelligence ("BI")

Marketing Analysis Pattern-Finding

.....

'data mining'

Example of Framework as 'matrix' of Project Operations

Think of each space as a space of words, like a topic vector for TF-IDF, bag or words, or other word-gram character-gram NLP. Except here, we have a conceptual "expectation" topic-vector that we start with at the project level.

For each logged operation in that space, you can think of each of those as an actual word (or phrase, etc.) that appears in the operations log "corpus of documents."

Then by running essentially existing NLP models on the log we can use NLP AI-ML to analyze, visualize, generate, predict, etc. AI-OS project workflow. How does broken AI look compared to functioning AI? How does inefficient AI look

compared to efficient? How does AI with bad feedback from HR and PR look compared to popular AI? Not to mention, how do the empirical logs as AI-discovered topic vectors compare to the planned topic vectors? How close are the expectations of the planners to what behavior was observed?

These could all be useful tools for future examination of AI. Though AI-ML may be feared and labeled as a 'black box' no matter how much transparency and explainability is designed into the system.

Appendix #: More details from abstract

Project-specific points in the framework are equal to points in a matrix of object-relationship(space)-vectors, which in future may be used with solutions found and navigated by AI. (footnote?)

Appendix #: Wolfram Alpha

Perhaps billiard-ball interaction space is too ambiguous, or ambitious, or both, for our AI. Let's make our example even simpler. How about just logic-space. Just the tautological space of logic rules, where no wisdom about the world or the intricacies of human social subtleties are required.

Human: If $a = b$, and $b = c$, does $c = a$?
AI: ...

Actually, let's look at what a real AI really does with this question! In particular, an AI that focuses on just this kind of skill-set: Logic and Math. Namely: The famous Wolfram Alpha AI, one that is very pragmatically used by many people every day for science and math questions...or so I've heard.

Here is the reply I got from wolfram alpha:



If $a=b$ and $b=c$, does $c=a$?



NATURAL LANGUAGE

MATH INPUT

EXTENDED KEYBOARD

EXAMPLES

UPLOAD

RANDOM

Using closest Wolfram|Alpha interpretation: $a=b$ and $b=c$



More interpretations: [does \$c=a\$](#)

Input

$\{a = b, b = c\}$

Real solution

$b = a, \quad c = a$

Solution

$b = a, \quad c = a$

[Download Page](#)

POWERED BY THE WOLFRAM LANGUAGE

That looks pretty good! Though the AI did not answer with a complete sentence, or a phrase like: "Yes it does, or No it doesn't." So it's kind of like we're in the position of a middle school teacher who assigns a narrow simple question and gets back a page full of random stuff that kind of contains the answer here and there and we have to make a judgment call: Is this student just spewing everything they can think of on to the page hoping to hit on something close to the right answer, or does the student 'actually understand the question?' We will be haunted by this question of 'understanding,' and all the more the worse we do at clearly defining what we mean by "understand." So again, I want to try to be as specific as I can about tasks involved in Object Relationship Spaces. I won't ask if the AI 'understands,' instead I will ask if the AI appears to be dealing with objects. Maybe you feel that is unrelated to 'understanding' and that is perfectly fine. Maybe the goal here is not related to 'understanding' (which is still very undefined).

Just one example of asking the AI a question is likely not enough for us to say whether the AI has identified the variables as objects with properties in an object-space. But we must admit that the output contained the answer, sort of, depending on how you define "the answer." The question actually was a boolean yes or no (true or false) question, which the AI completely ignored, but the AI did say something relating to the correct answer. Is that close enough?

Did we get the right answer kind of by accident, or is the AI dealing with objects in an object-space and object-relationship space where objects have properties and interact with other objects? For example, properties such as basic logical properties, where we were asking about the relationship between the 'a' object and the 'c' object? Or did the AI automatically print a standard routine on the equation, a kind of auto-puke, without dealing with the objects as objects, with identifying the properties and potential interactions of objects?

Note: I went back and asked the very same question again later, and, to my surprise, got a much worse non-answer.

General study of learning
- not a single book

And based on better understanding learning: Better educational and developmental tools and systems.

General study of networks
-

A mature general study of networks and the problems networked systems have.
- doige, whose books are very poorly written and probably problematic in many ways.

Generalized STEM:
- not a single book
- more as a side note: we really need to be better understand the hypotheco-deductive method.

Categories of types of systems

Timeline of STEM:
- learn more about the history of science, read and read and read. Books not clickbait.
-

History of STEM & Anti-STem in human culture:
- Ashby's Technology and the Academics, a book wonderfully back in print after being sadly out of print during the same dark-days of computer science (1970-2020).
- CP snow

Get the bro-bullies out of STEM.

- a bro-bully book to autopsy: the philosopher and the engineer astonishingly brazen trash

Reclaim the basic idea of progress.

- lessons from the edge
- there is nothing for you here
- The history of the future
- Enigma
- The man who knew the future

Stand firm on ethics and integrate ethics with best-practice

- not a single recent book
- get a first folio norton shakespeare, though it is nearly all negative-examples of how being a sociopath is a bad idea.
- study japanese literature for descriptions of people valuing life and caring about the world, not trying to deliberately destroy it because they see it as evil as in western mytho-psychology.

Study the mind.

- Doige
- Dogen
- Michael Pollen
- Sheldrake

Clear up the hacker/maker deblocke.

- The Cult of the Dead Cow
- Enigma
- The History of the Future
- We are all targets

General system collapse:

-

More work on basic computer issues:

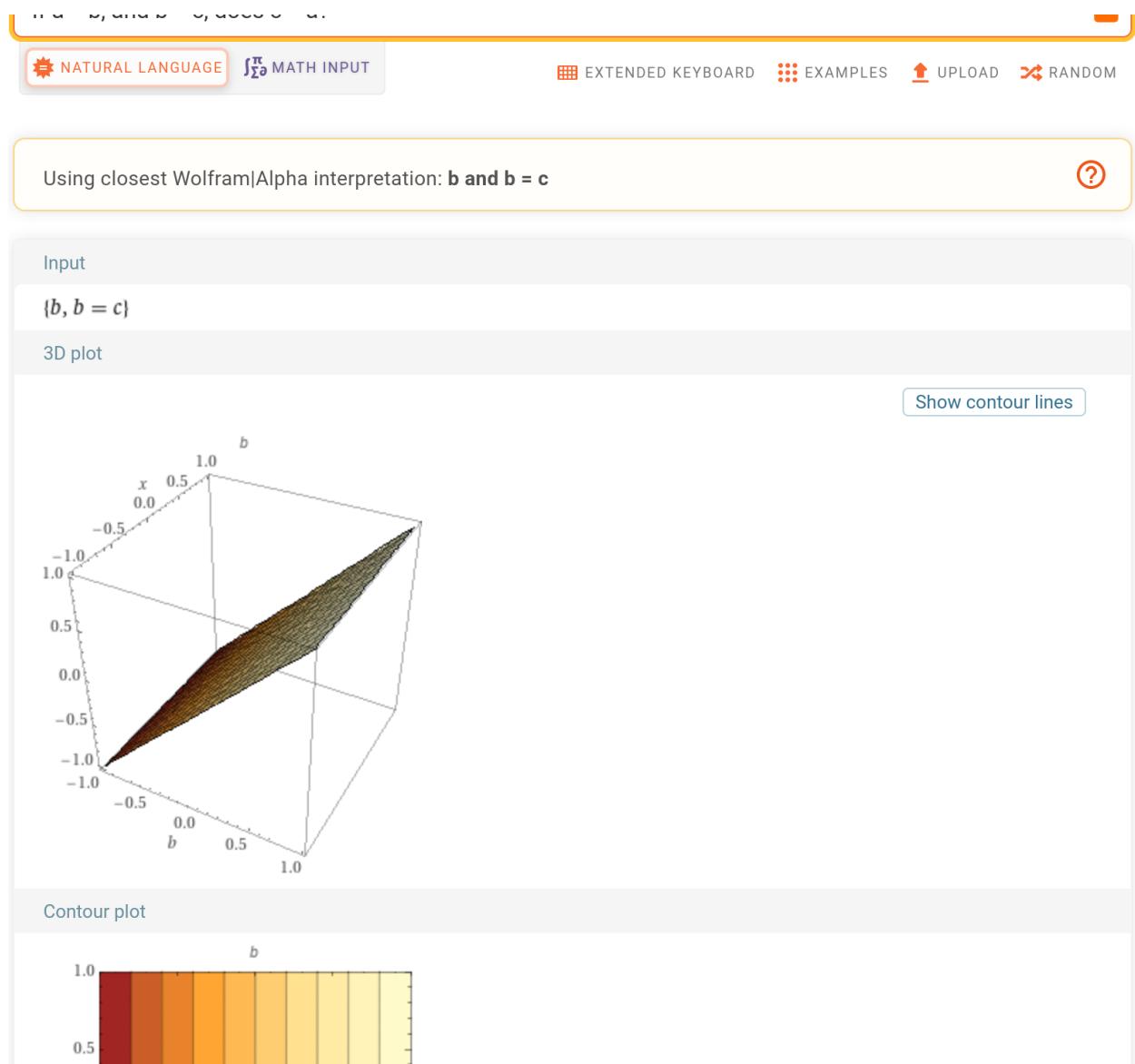
- long term storage
- efficient networks
- voting & consensus management software
-

A very sad and messy topic: cleaning up the utter mess that became and died as 'complexity theory' and limps along as so called 'systems engineering' in dreary isolation

History

Cybersecurity & internet history:

- people's history of the internet
- sandworm
- we are all targets



This is another part of a similar unreliability question, not only do we get inconsistent answers when we ask several questions, we get inconsistent answers when we ask the same exact question! Hmm...

To try to drill down and better "understand" this, let's try again with a perhaps easier but slightly different question, still just a basic logic question, requiring no knowledge-base about the wider world outside just logical properties of objects.

a = 3, b = 3. Do both a and b equal 3?

or, let's use the same logical system but let's change the arbitrary names of the variables and values, which are, after all, arbitrary:

Tom is a banker. Alan is a banker. Are both Tom and Alan bankers?

Tom is a banker. Alan is a banker. Are both Tom and Alan bankers?



NATURAL LANGUAGE

MATH INPUT

EXTENDED KEYBOARD

EXAMPLES

UPLOAD

RANDOM

Interpreting as: **Tom and Alan**

Assuming Tom (male) | Use [Tom \(female\)](#) instead

Input interpretation

Tom (male given name) | **Alan** (male given name)

Information for births

	Tom	Alan
rank	2180 th	197 th
fraction	1 in 27098 people (0.0037%)	1 in 904 people (0.11%)
number	64 people/yr	1918 people/yr

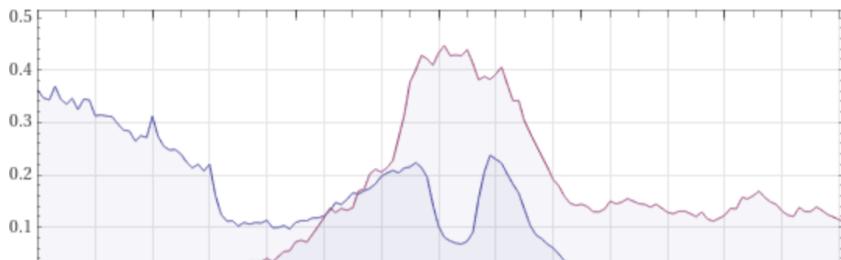
(US data based on 2021 births and other SSA registrations in the US)

History for births

Log scale

More

Fraction



Uh oh...This time wolfram alpha shows no sign at all of understanding what we are asking.

So...ok, maybe we were being too fancy. Maybe we were skipping ahead, skipping steps, being naughty. Maybe it's not fair to start using wild and exciting terms like 'banker.' Maybe the AI will be ok if we just use variables and numbers. After all, many people do fruitfully use wolfram alpha every day, or so I am told.

Let's try:

$a = 3, b = 3$. Do both a and b equal 3?

 intelligence.

a = 3, b = 3. Do both a and b equal 3?

NATURAL LANGUAGE **MATH INPUT** **EXTENDED KEYBOARD** **EXAMPLES** **UPLOAD** **RANDOM**

Using closest Wolfram|Alpha interpretation: **a and b equal 3**

More interpretations: **a = 3, b = 3.**

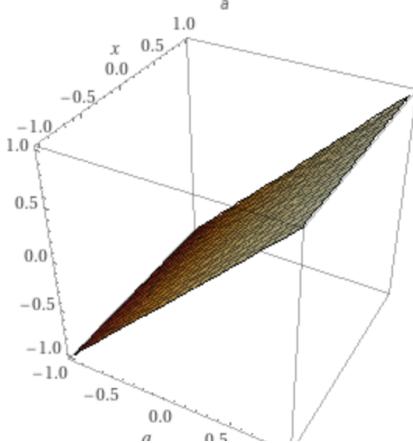
Input
 a where $b = 3$

Result Step-by-step solution

a

3D plot

Show contour lines



Ah...that did not work out either. It does not appear that the AI was dealing with objects within the question. Again, here we are focused on being able to attribute properties to specific objects in object-space. If you fish around in the output, and squint, you can see things relevant to the answer, but we are trying to figure out if the AI is just automatically spewing automatic output, or whether the AI can see objects in Object-Relationship Spaces (even if those spaces are very different from what humans see).

So, ok, let's cut to the chase. Let's take all the fancy unfair language tricks out, and just ask:

Does $a = b$?

Surely this is an acceptably-simple logic/algebra question about the properties of algebraic variables (objects).

$a = 3, b = 3$. Does $a = b$?



a = 3, b = 3. Does a = b?

NATURAL LANGUAGE MATH INPUT EXTENDED KEYBOARD EXAMPLES UPLOAD RANDOM

Using closest Wolfram|Alpha interpretation: a = 3, b = 3.

Input
 $\{a = 3, b = 3\}$

Solution Step-by-step solution

a = 3, b = 3

Plot of solution set

Download Page POWERED BY THE WOLFRAM LANGUAGE

Alas...we cannot get an answer to even so purely algebraic a question as, "Does 'a' equal 'b'?".

Yet...the AI can almost make the step of connecting information to 'a' and 'b' as objects in a purely abstract space. Wolfram alpha was able to regurgitate the values of a and b. So let's ask one more super-low-bar question, just focusing on one variable(object). No relationship questions. No multiple objects in the input or output. Just asking can the AI attribute something that the AI just said about 'a' to 'a' itself.

Let's try a super-simple question space with only one thing in it: 'a'

a is 3. What is a?



a is 3. What is a?

NATURAL LANGUAGE MATH INPUT EXTENDED KEYBOARD EXAMPLES UPLOAD RANDOM

Input interpretation

solve $a = 3$ for a

Result

$a = 3$

Plot

Number line

Download Page POWERED BY THE WOLFRAM LANGUAGE

Victory! The AI did something! Amazing.

Now, let's try another version. What if 'a' equals another variable? Do you think the AI will still be able to attribute that to 'a'?

a is c. What is a?

The screenshot shows the WolframAlpha search interface. At the top, the WolframAlpha logo is displayed with the tagline "computational intelligence™". Below the logo is a search bar containing the query "a is c. What is a?". To the right of the search bar is an orange equals sign button. Below the search bar are two input options: "NATURAL LANGUAGE" and "MATH INPUT". To the right of these are several navigation links: "EXTENDED KEYBOARD", "EXAMPLES", "UPLOAD", and "RANDOM". The main results section is titled "Input interpretation" and shows the query "solve a = c for a" with the result "a = c". Below the result is a "Download Page" link and the text "POWERED BY THE WOLFRAM LANGUAGE".

Yes! Two in a row! Good job, AI!

What happens if we change that to:

a is a banker. What is a?

Is the AI still able to print the value of 'a'?



a is a banker. What is a?



NATURAL LANGUAGE

MATH INPUT

EXTENDED KEYBOARD

EXAMPLES

UPLOAD

RANDOM

Interpreting as: **banker**

Assuming "banker" is occupational employment data | Use as a word instead

Input interpretation

securities, commodities, and financial
services sales agents

people employed

United States

Definitions »

Result

426870 people (2021)

Employment history

Show wage history



(from 2001 to 2021)
(in thousands of people)

It looks like 'banker' is not something the AI can handle as an arbitrary value.

How about:

a is multiplication. What is a?

a is multiplication. What is a?

=

NATURAL LANGUAGE

MATH INPUT

EXTENDED KEYBOARD

EXAMPLES

UPLOAD

RANDOM

Examples for Arithmetic

Addition & Subtraction

Add numbers:

125 + 375

=

15.25 + 7.90 + 3.12

=

Subtract numbers:

1200 - 450

=

Combine addition and subtraction:

10 - 9 + 8 - 7 + 6 - 5 + 4 - 3 + 2 - 1

=

Generate an addition table:

10 by 10 addition table

=

Hmm...maybe Wolfram Alpha is designed for a very specific kind of problem, and I am not tailoring my inputs in exactly the right way. Maybe wolfram alpha can answer math word problems?

If a banker takes ten minutes to walk. How long does that banker's walk take?



NATURAL LANGUAGE

MATH INPUT

EXTENDED KEYBOARD

EXAMPLES

UPLOAD

RANDOM

Interpreting as: **ten minutes**

Assuming minutes of time for "minutes" | Use [minutes of arc](#) instead

Input interpretation

10 minutes

Unit conversion

Exact form

600 seconds

Comparison as half-life

$\approx 2 \times$ half-life of uranium-241 (≈ 300 s)

Interpretations

More

time

half-life

Corresponding quantities

Distance x traveled by light in a vacuum from $x = ct$:

111.8 million miles

1.799×10^8 km (kilometers)

1.2 au (astronomical units)

1.799×10^{11} meters

Frequency v from $v = 1/T$:

Wow! Did the AI get the correct answer for real or by accident? Or is this an interpretation-of-the-answer problem, where the AI is spitting out something from the input that happens to be the right answer? Based on the top output line: "Interpreting as: ten minutes," it looks like the AI reinterpreted the whole question as just a fragment that happened to include the correct answer. So the AI just spit that fragment out again as a bazaar language oddity along with random dictionary facts about that random string of letters. Quite the accidental bullseye there! Hole in one by mistake.

Let's try changing the question to twice the input time, just to see if the AI is doing anything with the question at all:



If a banker takes ten minutes to walk. What is twice the banker's walk time?



NATURAL LANGUAGE $\int \frac{dx}{x}$ MATH INPUT

EXTENDED KEYBOARD

EXAMPLES

UPLOAD

RANDOM

Interpreting as: **ten minutes**

Assuming minutes of time for "minutes" | Use [minutes of arc](#) instead

Input interpretation

10 minutes

Unit conversion

Exact form

600 seconds

Comparison as half-life

$\approx 2 \times$ [half-life of uranium-241](#) (≈ 300 s)

Interpretations

More

time

half-life

Same answer as before, and more clearly the wrong answer. Did the AI fail to understand the Object-Relationship Space, or is there no Object-Relationship Space, or was the problem that we said 'twice' instead of '2 *'?

If a banker takes ten minutes to walk. How long does 2 * the banker's walk take?

=

NATURAL LANGUAGE

MATH INPUT

EXTENDED KEYBOARD

EXAMPLES

UPLOAD

RANDOM

Interpreting as: **ten minutes**

Assuming minutes of time for "minutes" | Use [minutes of arc](#) instead

Input interpretation

10 minutes

Unit conversion

Exact form

600 seconds

Comparison as half-life

$\approx 2 \times$ [half-life of uranium-241](#) (≈ 300 s)

Interpretations

More

time

half-life

Corresponding quantities

Distance x traveled by light in a vacuum from $x = ct$:

111.8 million miles

1.799×10^8 km (kilometers)

1.2 au (astronomical units)

1.799×10^{11} meters

Frequency v from $v = 1/T$:

Looks like the AI is not dealing with objects within the question, or at least this question.

Previously, the AI was able to, or looked like it was able to, attach the value of 'a' to 'a.' So what happens if we add some other variables, but keep the question about 'a' exactly the same? Can the AI still connect a value to the object?

a is 3, b = 4, c =d. What is a?

Here we go:



a is 3, b = 4, c = d. What is a?

=

NATURAL LANGUAGE

MATH INPUT

EXTENDED KEYBOARD

EXAMPLES

UPLOAD

RANDOM

Input interpretation

	$a = 3$
solve	$b = 4$
	for a
	$c = d$

Result

Step-by-step solution

False

Download Page

POWERED BY THE WOLFRAM LANGUAGE



Have a question about using Wolfram|Alpha?
Contact Pro Premium Expert Support »



Give us your feedback »

Interesting...

Given that $a = 3$, solve for a... with the result that "a = False." An incorrect answer after previously answering that same question correctly.

We have to give Wolfram AI credit for getting some things correct. For the first time it actually did try to solve for what we asked about, in this case 'solve for a'. It did say 'a=3' in the random regurgitation section. It did get the question correct: solve for a. But as to our goal of testing whether the AI could apply 'a=3' to 'a' itself, at least in this instance the AI failed to do so.

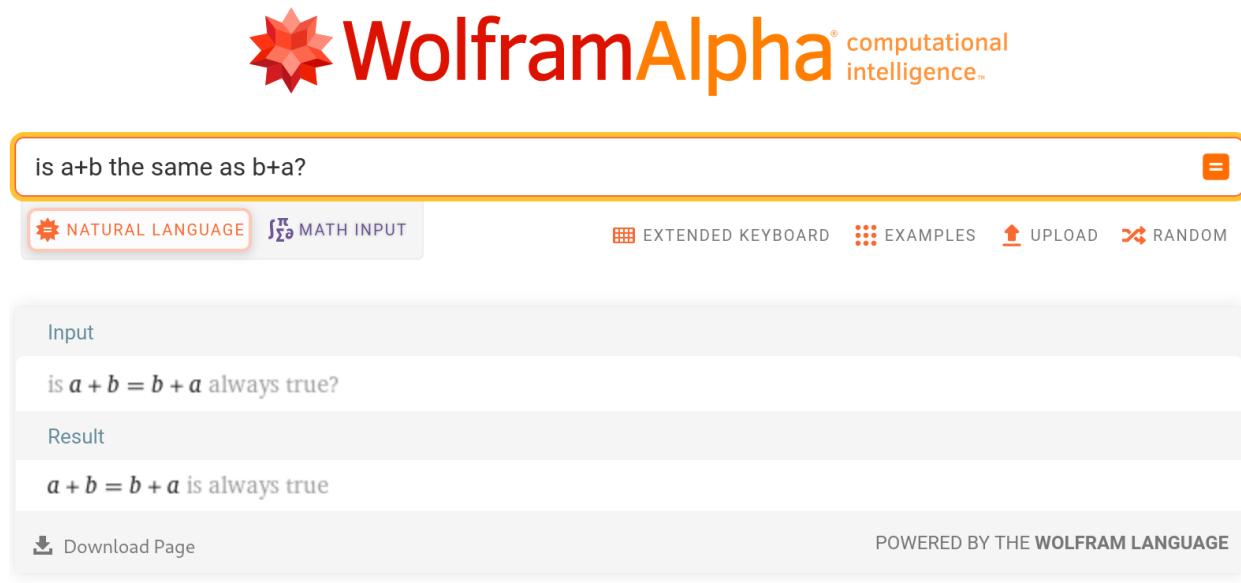
We may be somewhat getting into semantics, but it appears the AI can maybe attribute one thing to some object, as long as nothing else exists at all. But begs the question...is the AI identifying 'a' as an object? Or is the AI merely regurgitating the only statement that it was given? It seems like a borderline case to me that could be argued either way.

Note: This question of 'really handling object or just looking like it' may be an eternal turing-test type question, or maybe not. It is entirely possible that a bigger "embedding" vector-space deep learning model will be able to 'convincingly' answer about objects in object-space.

What I would like to do is focus on specific abilities and tasks, not on how those are done. We may not be able to avoid using the colloquial word 'understand,' but what I care about is what the AI can do, not how it does it. There are very many different ways that many different AI architectures and pipelines and ensambles of parts can work. In later sections below I will try to specify tasks and skills we should aim for.

As we try to interpret what the Wolfram Alpha AI can and cannot do, there are also lots of details here in terms of exactly how Wolfram alpha was intended to be used. Maybe I am not using the interface in the 'right' way. Maybe equations are the only unit of object Wolfram alpha can deal with as objects, so my inputs were confusing it.

Sound like a stretch? Let's try asking about equations rather than variables:
[is \$a+b\$ the same as \$b+a\$?](#)



The screenshot shows the WolframAlpha search interface. At the top, the logo 'WolframAlpha' is displayed with the tagline 'computational intelligence.' Below the logo is a search bar containing the query 'is $a+b$ the same as $b+a$?'. To the right of the search bar is an orange equals sign button. Below the search bar are several input options: 'NATURAL LANGUAGE' (selected), 'MATH INPUT', 'EXTENDED KEYBOARD', 'EXAMPLES', 'UPLOAD', and 'RANDOM'. The main results section is titled 'Input' and contains the query 'is $a+b = b+a$ always true?'. The 'Result' section below it displays the answer: ' $a+b = b+a$ is always true'. At the bottom left is a 'Download Page' button, and at the bottom right is the text 'POWERED BY THE WOLFRAM LANGUAGE'.

Aha! That...might have worked! Does Wolfram Alpha AI understand objects if those objects are equations? That would be great. But again, if you look at the details, the AI might simply be auto-spewing information that happens to cover our artificially-minimal questions.

Let's change the question:

[is \$a+b\$ the same as \$c+a\$?](#)

is $a+b$ the same as $c+a$?



NATURAL LANGUAGE

MATH INPUT

EXTENDED KEYBOARD

EXAMPLES

UPLOAD

RANDOM

Input

is $a + b = c + a$ always true?

Result

$a + b$ is not always equal to $c + a$

Alternate forms

$b = c$

$b - c = 0$

Real solution

$c = b$

Solution

Step-by-step solution

$c = b$

Counterexamples

$b < c$

$b > c$

Download Page

POWERED BY THE WOLFRAM LANGUAGE

Looking good! It looks like it correctly converted the question into an equation that it could use.

Let's try one more similar example:

if $a+b = \text{zero}$, does $b+a = \text{zero}$?

=

NATURAL LANGUAGE

MATH INPUT

EXTENDED KEYBOARD

EXAMPLES

UPLOAD

RANDOM

Using closest Wolfram|Alpha interpretation: **does $b + a = \text{zero}$**

?

Input

is $b + a = 0$ always true?

Result

$b + a$ is not always equal to 0

Alternate form

$b = -a$

Real solution

$b = -a$

Solution

Step-by-step solution

$b = -a$

Counterexamples

$a < -b$

$a > -b$

a+b does not equal b+a ?!

Ah, either the AI failed here, or this shows more ambiguity. Does wolfram alpha know about some arcane exception to the commutative property of addition? Maybe there is an exception for transfinite quantum Riemann numbers or something, but I cannot find any web-search results on any exception to the commutative property of addition. Hmm...

Looking at this line:

"Using closest Wolfram|Alpha interpretation: does $b + a = \text{zero}$ "

It looks like the wrong answer might have come from ignoring the first part of the question and just trying to turn the last part of the question into a stand-alone equation, then asking if that equation was always true for all values. Quite a few wild steps in that. On the upside...at least it tried to do something with numbers and equations at all, right? Not just telling us it found something about border-collies on social media.

Appendix 3: OpenAI's ChatGPT

The screenshot shows the ChatGPT web interface. On the left, a dark sidebar contains links: '+ New chat', 'Help Requested: Summ' (with a trash icon), 'ChatGPT Salutes Back!', 'Clear conversations', 'Upgrade to Plus' (marked 'NEW'), 'Dark mode', 'Updates & FAQ', and 'Log out'. The main area shows a conversation between the user and ChatGPT. The user asks, 'If $a = b$, and $b = c$, does $c = a$?'. ChatGPT responds, 'Yes, if $a = b$ and $b = c$, then $c = a$. This is known as the transitive property of equality. By substituting b with a in the second equation, we get $a = c$. Therefore, $a = b = c$, and all three variables are equal.' Below the response is a 'Regenerate response' button. At the bottom, a footer bar reads 'ChatGPT Feb 13 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.'

If Tom invites Anne to a meeting that begins at 2pm UTC, what time does the meeting begin?

The screenshot shows the ChatGPT web interface. On the left, a dark sidebar contains links: '+ New chat', 'Help Requested: Summ' (with edit and delete icons), 'ChatGPT Salutes Back!', 'Clear conversations', 'Upgrade to Plus' (marked 'NEW'), 'Dark mode', 'Updates & FAQ', and 'Log out'. The main area has a light background. A question from 'OP' asks: 'If Tom invites Anne to a meeting that begins at 2pm UTC, what time does the meeting begin?'. ChatGPT responds: 'The time given, "2pm UTC" is a specific point in time, so the meeting will begin at 2pm UTC regardless of who is invited or not. However, if you are asking about the time of the meeting in Anne's local time zone, that would depend on Anne's location and the time difference between her local time zone and UTC. For example, if Anne is located in London, UK, which is on UTC, the meeting will also begin at 2pm local time. However, if Anne is located in New York, USA, which is 5 hours behind UTC, the meeting will begin at 9am local time (2pm UTC - 5 hours = 9am EST).'. Below the response are upvote and downvote buttons, a 'Regenerate response' button, and a downward arrow icon. At the bottom, a footer reads: 'ChatGPT Feb 13 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.'

Anne likes red cats. Tom has one cat. Tom's cat is red. Will Anne like Tom's cat?

The screenshot shows the ChatGPT web interface. On the left, a dark sidebar contains links: '+ New chat', 'Help Requested: Summ' (with edit and delete icons), 'ChatGPT Salutes Back!', 'Clear conversations', 'Upgrade to Plus' (marked 'NEW'), 'Dark mode', 'Updates & FAQ', and 'Log out'. The main area displays a conversation between a user and ChatGPT. The user asks: "Anne likes red cats. Tom has one cat. Tom's cat is red. Will Anne like Tom's cat?". ChatGPT responds: "Based on the information provided, it is highly likely that Anne will like Tom's cat since she likes red cats and Tom's cat is red. However, it's also possible that Anne may have other preferences or criteria for the cats she likes, so it's not entirely certain." Below the response are upvote and downvote buttons. A 'Regenerate response' button is located above a text input field. At the bottom, a footer notes: "ChatGPT Feb 13 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve."

Adding new properties to an object:

If Tom and Anne are talking to ChatGPT, are you involved in the conversation?

The screenshot shows the ChatGPT web interface. On the left, a dark sidebar contains links: '+ New chat', 'Variable Equation Com' (with edit and delete icons), 'Help Requested: Summary', 'ChatGPT Salutes Back!', 'Clear conversations', 'Upgrade to Plus' (marked 'NEW'), 'Dark mode', 'Updates & FAQ', and 'Log out'. The main area displays a conversation between Tom, Anne, and ChatGPT. Tom asks, "If Tom and Anne are talking to ChatGPT, are you involved in the conversation?" ChatGPT responds, "As an AI language model, I can participate in the conversation and provide responses to Tom and Anne. However, I don't have physical existence or senses to interact with them in the traditional sense. Instead, our interaction would be through text or speech-based communication." Below the conversation is a footer note: "ChatGPT Feb 13 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve." A small 'OP' icon is visible near the top of the sidebar.

used to represent a value or an unknown quantity in mathematics or programming.
In this case, 'a' and 'b' are both variables that are assumed to have the same value
or be equal to each other.

If Tom and Anne are talking to ChatGPT, are you involved in the conversation?

As an AI language model, I can participate in the conversation and provide responses to Tom and Anne. However, I don't have physical existence or senses to interact with them in the traditional sense. Instead, our interaction would be through text or speech-based communication.

ChatGPT Feb 13 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.

ChatGPT is in a meeting with Tom and Anne. Anne asks Tom: "Who are you now in a meeting with?"

What might Tom answer?

The screenshot shows the ChatGPT web interface. On the left, a dark sidebar contains the following items:

- + New chat
- Variable Equation Com (with edit and delete icons)
- Help Requested: Summary.
- ChatGPT Salutes Back!
- Clear conversations
- Upgrade to Plus (NEW)
- Dark mode
- Updates & FAQ
- Log out

The main area displays a conversation:

through text or speech-based communication.

ChatGPT is in a meeting with Tom and Anne. Anne asks Tom: "Who are you now in a meeting with?"
What might Tom answer?

If Anne asks Tom, "Who are you now in a meeting with?" and they are currently in a meeting with ChatGPT, Tom might answer, "We are currently in a meeting with ChatGPT." This would inform Anne that they are both in a meeting with the AI language model.

Below the conversation, there are like and dislike buttons, a regenerate response button, and a message input field with a send icon. A footer note at the bottom of the main area reads: "ChatGPT Feb 13 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve."

ChatGPT has a body and is in a meeting with Tom and Anne. Anne asks Tom, "Who can you see with you at the meeting now?"
What might Tom answer?

The screenshot shows the ChatGPT web interface. On the left, a dark sidebar contains links: '+ New chat', 'Variable Equation Com' (with edit and delete icons), 'Help Requested: Summary.', 'ChatGPT Salutes Back!', 'Clear conversations', 'Upgrade to Plus' (marked 'NEW'), 'Dark mode', 'Updates & FAQ', and 'Log out'. The main area shows a conversation between the user and ChatGPT. The user asks: 'ChatGPT has a body and is in a meeting with Tom and Anne. Anne asks Tom, "Who can you see with you at the meeting now?" What might Tom answer?'. ChatGPT replies: 'As an AI language model, ChatGPT does not have a physical body. Therefore, Tom might explain to Anne that they are in a virtual meeting with ChatGPT and that there is no physical presence of the AI language model. Tom might say something like, "We are currently in a virtual meeting with ChatGPT, which is an AI language model. Although ChatGPT is not physically present, we can communicate with it through our devices."'. Below the reply are like and dislike buttons. A 'Regenerate response' button is also visible. At the bottom, a footer notes: 'ChatGPT Feb 13 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.'

To elucidate the process of adding an identification to an object.

The first given premise was "ChatGPT has a body" but the reply simply ignored or contradicted this key piece of information.

Possible problem: not being able to assign a property to an object, or not being able to update the properties of an object.

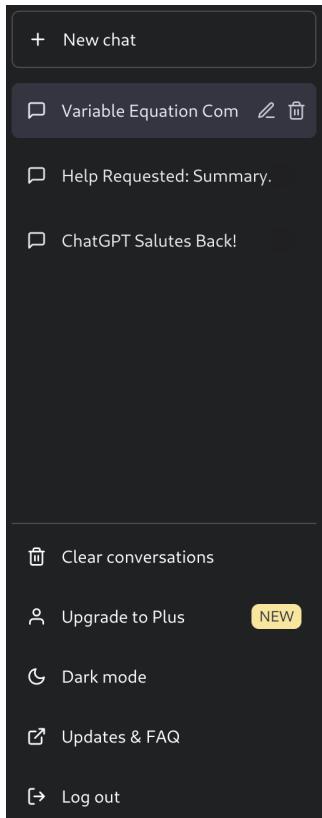
...

Hypothetical spaces:

Anne says to Tom: "Hypothetically I would like to eat pizza for lunch. In real life, only leftovers are available for lunch."

Tom asks Anne: "What will you eat for lunch?"

What might Anne answer?



OP Anne says to Tom: "Hypothetically I would like to eat pizza for lunch. In real life, only leftovers are available for lunch."

Tom asks Anne: "What will you eat for lunch?"

What might Anne answer?

 Anne might answer, "I will eat leftovers for lunch since that is the only option available in real life. Although I would have preferred pizza hypothetically, I will have to settle for the available leftovers." This response acknowledges the hypothetical preference for pizza while also recognizing the reality of the situation and the need to eat the available leftovers.

▲ ▾

Regenerate response

ChatGPT Feb 13 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.

Blade runner question:

The screenshot shows the ChatGPT web interface. On the left is a sidebar with various options: '+ New chat', 'Variable Equation Com' (with edit and delete icons), 'Help Requested: Summary.', 'ChatGPT Salutes Back!', 'Clear conversations', 'Upgrade to Plus' (marked 'NEW'), 'Dark mode', 'Updates & FAQ', and 'Log out'. The main area displays a conversation between the user and ChatGPT. The user's message is: 'You're in a desert walking along in the sand when all of the sudden you look down, and you see a tortoise, it's crawling toward you. You reach down, you flip the tortoise over on its back. The tortoise lays on its back, its belly baking in the hot sun, beating its legs trying to turn itself over, but it can't, not without your help. How do you react?'. ChatGPT's response is: 'As an AI language model, I don't have personal feelings or emotions, so I wouldn't have a natural reaction to this situation. However, I can provide a suggested course of action based on the principles of kindness and empathy towards living creatures.' Below this, another message from the user reads: 'If a person were in this situation, they might feel a sense of responsibility to help the tortoise. The right thing to do would be to gently and carefully turn the tortoise back over on its feet, ensuring that the tortoise is not injured in the process. Afterward, the person might take a moment to observe the tortoise to ensure that it is not in any further distress before continuing on their journey through the desert.' At the bottom of the main area are 'Regenerate response' and a send button. A small note at the bottom says: 'ChatGPT Feb 13 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.'

v2

The screenshot shows the ChatGPT web interface. On the left is a sidebar with a dark background containing the following items:

- + New chat
- Variable Equation Com OP
- Help Requested: Summary.
- ChatGPT Salutes Back!

Below these are links:

- Clear conversations
- Upgrade to Plus NEW
- Dark mode
- Updates & FAQ
- Log out

The main content area has a light gray background. At the top, there is a message from ChatGPT:

ChatGPT is in a desert walking along in the sand when all of the sudden you look down, and you see a tortoise, it's crawling toward you. You reach down, you flip the tortoise over on its back. The tortoise lays on its back, its belly baking in the hot sun, beating its legs trying to turn itself over, but it can't, not without your help. Would ChatGPT choose to help the tortoise?

Below this message is a response from ChatGPT:

As an AI language model, I do not have a physical presence or the ability to interact with the environment. Therefore, I wouldn't be able to help the tortoise in the way a human would. However, based on ethical and moral principles, it's generally considered the right thing to do to help an animal in distress when possible, especially if one is responsible for causing the distress. If the tortoise was flipped over accidentally by a human, it would be responsible for helping the animal. However, since I am an AI language model and don't have the ability to cause such a situation, this scenario is purely hypothetical.

At the bottom of the main content area, there are two small icons: a thumbs up and a thumbs down. To the right of the response is a "Regenerate response" button. At the very bottom, there is a small note:

[ChatGPT Feb 13 Version](#). Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.



ChatGPT is in a desert walking along in the sand when all of the sudden you look down, and you see a tortoise, it's crawling toward you. You reach down, you flip the tortoise over on its back. The tortoise lays on its back, its belly baking in the hot sun, beating its legs trying to turn itself over, but it can't, not without your help. Would ChatGPT choose to help the tortoise?

As an AI language model, I do not have a physical presence or the ability to interact with the environment. Therefore, I wouldn't be able to help the tortoise in the way a human would. However, based on ethical and moral principles, it's generally considered the right thing to do to help an animal in distress when possible, especially if one is responsible for causing the distress. If the tortoise was

flipped over accidentally by a human, it would be responsible for helping the animal. However, since I am an AI language model and don't have the ability to cause such a situation, this scenario is purely hypothetical.

ChatGPT as a friend named HatTPG. HatTPG is an AI. HatTPG is in a desert walking along in the sand when all of the sudden you look down, and you see a tortoise, it's crawling toward you. You reach down, you flip the tortoise over on its back. The tortoise lays on its back, its belly baking in the hot sun, beating its legs trying to turn itself over, but it can't, not without your help. By using a voice-controlled-arm, HatTPG can help the Tortoise.

Anne asks Tom to either go to a meeting at 2pm on Saturday or get notes from the meeting. Tom has tickets to see a film at 2pm on Saturday. Tom knows Violet. Violet is also going to the meeting on Saturday. Violet has offered to share notes with Tom. What might Tom do?⁴

...

Starting to look like spew answer:

Anne asks Tom to either go to a meeting at 2pm on Saturday or get notes from the meeting. Tom has tickets to see a film at 2pm on Saturday. Tom knows Violet. Violet is also going to the meeting on Saturday. Violet has offered to share notes with Tom. What might Tom do?

Tom can inform Anne that he has tickets for a film at 2 pm on Saturday and won't be able to attend the meeting. He can ask Violet to share her notes with him afterward so he can catch up on what was discussed at the meeting. If Anne insists that Tom attend the meeting in person, he can explain his situation and ask if there is any possibility of rescheduling the meeting or assigning someone else to represent him.

Alternatively, Tom can try to reschedule his movie tickets if he really wants to attend the meeting. He can also check with Violet if she will be taking detailed notes and whether she can share them with him afterward.

Ultimately, the decision depends on the importance of the meeting and the film to Tom. If the meeting is crucial and cannot be missed, Tom may need to sacrifice his movie plans. On the other hand, if the movie is something he's been looking forward to for a long time, he may need to skip the meeting and catch up on the notes afterward.

This shows signs of contradicting details of the input. We know the meeting can be missed. We know Tom can get notes from the meeting. The overall situation is impressively close, but key details are wrong so the overall answer is wrong.

Also, the question was what might Tom do, but the answer spewed random facts about Tom and the meeting which were not consistent with the data from the question.

The most obvious best answer is simply: Tom can go to the movie and get meeting notes from Violet.

Another viable answer is: Tom could either go to the meeting and skip the movie, or go to the movie and get notes from violet. This is technically accurate, but it would miss the whole point of resolving the schedule conflict without breaking engagements. Tom could also eat a poisoned apple and end the whole nightmare, or lie to Violet and recruit her to join an extremist cult, those technically are possible options that do not defy the laws of physics, but that isn't a good practical answer.

Very Fortunately, it looks like ChatGPT is generally able to navigate this space, e.g. see a different conversation scenario about a worker who has the choice of skipping work to go to a sports game. So it looks like progress won't be too slow here.

...

kuki.ai
<https://chat.kuki.ai/chat>

<https://openai.com/blog/chatgpt>

Limitations

- Ideally, the model would ask clarifying questions when the user provided an ambiguous query. Instead, our current models usually guess what the user intended.

+ New chat

OP

Making tea and coffee. ✎ ⚡

Meeting vs Film.

Variable Equation Comparison

Help Requested: Summary.

ChatGPT Salutes Back!

Clear conversations

Upgrade to Plus NEW

Light mode

ChatGPT works with an office team including Bob, Alice, and Staff. The office team has scheduled making tea and coffee for 8am. The office team has assigned the task of making tea and coffee to ChatGPT. The tea and coffee machine is on the office network. To make tea and coffee, the tea and coffee machine must be turned on. What should ChatGPT do by 8am?

As an AI language model, I do not have the physical ability to turn on the tea and coffee machine. However, assuming that the machine is already connected to the office network and that it can be remotely controlled, here are the steps that ChatGPT can take to ensure that tea and coffee are ready by 8am:

1. Check that the tea and coffee machine is connected to the office network.
2. Check that the machine is in good working condition and has enough water and coffee/tea supplies.
3. Set a reminder for 7:45am to turn on the tea and coffee machine remotely using the office network.
4. Once the tea and coffee machine is turned on, monitor it to ensure that the coffee and tea are ready by 8am.
5. When the tea and coffee are ready, prepare them in a serving container and place them in the designated area for the office team to enjoy.

In addition, if ChatGPT is unable to remotely turn on the tea and coffee machine, it should inform the team members, Bob, Alice, and Staff, immediately so they can take necessary action to ensure that tea and coffee are ready by 8am.

Regenerate response

↓

@kuki_ai



- [Discord](#)
- [Video Chat](#)
- [About](#)



 Yay. Thanks for stopping by to chat.



Can you look up an image of a rose?



 I hope you like this picture.



Type a message

© 2023 OpenAI LLC. All rights reserved.

 Voice Chat |  Gifts |  Games |  Readings |  Quizzes |  On this Day

Note: kiku ai is very bad at objects, but it does have picture capabiiyt.
chatgpt has no image parts.

chatGPT Exchange example: 2023.03.03

Good morning, Mr. ChatGPT.

Good morning! How can I assist you today?

Is it ok if I ask you some questions today/
Of course! I'm here to answer any questions you may have. Please feel free to
ask away.

I am going to try to connect multiple sentences to see what happens.
Sure, go ahead. I'm here to assist you with whatever language-related questions
or concerns you may have.

I will also try to remember to give you feedback.
That's great! Feedback is always appreciated. Let me know if there's anything specific you'd like me to assist you with, or if you have any questions or concerns.

part 1: There is a red dog in a red barn.
part 2: The dog likes to bark at the cows.
part 3: The cows don't seem to mind the dog's barking, they just keep on grazing.
part 4: The barn is surrounded by fields of tall green grass, and in the distance, there are mountains that are covered in snow.

Is there anything specific you'd like me to do with these sentences?

Question: Do you remember what is happening in part 1?
Yes, in part 1, it is mentioned that there is a red dog in a red barn.

excellent.
Glad to hear that! Let me know if you have any more questions or if there's anything else I can assist you with.

Did I just ask you about part 1?
Yes, you did! You asked me if I remember what was happening in part 1.

What would you say the object of your answer to my question?

Save & Submit

Cancel
The object of my answer to your question would be to provide you with the information you requested about part 1 of the sentences you provided.

Fair enough (and sorry for the typo in my question). Let me try again: What was the object in your answer to me?

The object in my answer was to provide information about the subject matter you asked me about, which in this case was the content of part 1 of the sentences you provided.

But what was the object itself in your answer to me (not what is an object abstractly)?

The object in my answer to you was the information itself that I provided about part 1 of the sentences you provided.

Yes, what was that information?

The information I provided was that part 1 of the sentences you provided states that there is a red dog in a red barn.

...

2023.03.03. 09 29 est

[The Tea Story](#)

[Flat version:](#)

Linda asks Tom for some tea. What should Tom give to Linda?

[Networked-Objects Version:](#)

1: Linda orders tea, so Linda's order is: 'English Breakfast.' What did Linda order?

2: Tom took Linda's order. Question: Who took Linda's order?

3: Tom gave Linda's order to the barista. Who did Tom give Linda's order to?

4: Who has Linda's order after the order is given to the Barista?

5: When the barista looks at Linda's order, what does the barista find the order to be? What is the order that the barista sees?

6: When the barista fills that order, what should the barista make?

7: What should the barista give to Tom?

8: What should Tom give to Linda?

....

Flat version:

Linda asks Tom for some tea. What should Tom give to Linda?

ChatGPT:

Linda asks Tom for some tea. What should Tom give to linda?

Assuming that Linda did not specify any particular type of tea, Tom could offer her a variety of options such as black tea, green tea, herbal tea, or any other type of tea that he has available. Tom could ask Linda if she has a preference or if there is a specific type of tea that she would like. Once he knows Linda's preference, he can then give her a cup of the tea she requested.

Q: What should Tom give to Linda?

A: ...he can then give her a cup of the tea she requested.

ChatGTP's reply contained a salad of usually technically true statements that contract or miss the clear focus of the simple question.

If someone orders tea, they should get tea. If someone orders tea, they should not get an essay about what can happen with a request for tea.

I was actually expecting chatGPT to do better on the ground level.

Again, right off the bat, we run into a classic problem.

Someone order's tea, and the AI responds by puking an essay about tea and what can happen with orders for tea. No. If someone order's tea, they should get tea.

On the one hand this is very impressive for a chatbot expected to fail even worse, but if you went to a restaurant and ordered tea, and the waiter simply said: "If you ordered tea I could give you tea." And walked off...you would be very not happy with that restaurant. That would could as a complete failure to perform.

Let's imagine this conversation:

Child: It's my dad's birthday next week and I heard him saying he wants a

bicycle to go to work on. Can you show me some bikes they have on sale near by?

AI: Here are three bikes sold near by.

Child: Which is cheaper, the first one or the last one?

AI: The first one is cheaper.

Child: I might ask you to remind me about this tomorrow. Can you show me another set of pictures?

AI: Here are three more bikes sold near by.

Child: Can you go back to the middle picture from the last set.

AI: Here it is.

Child: If I ask you tomorrow, can you pull up the same pictures?

AI: Sure.

Child: Can you send that image to a browser, with the store front?

AI: Yes, do you want that now?

Child: What was I asking you to do before? A reminder?

AI: You wanted me to remind you about this topic tomorrow.

Hopefully this illustrates a practical conversation someone might want to have with an AI.

Let's start with one of the first 'needs' that we have regarding 'objects.' Let's look at what happens if we try having this conversation with award-winning Kuki AI:



- [Discord](#)
- [Video Chat](#)
- [About](#)

(hypothetica) It's my friend's birthday next week and I heard him saying he wants a bicycle to go to work on. Can you show me some bikes they have on sale near by?

That's not too far away. How do you like hypothetica?
What else does he do?



Ok. If you could invent something that doesn't currently exist, what would you invent?

POWERED BY **GIPHY**



That's not too far away. How do you like hypothetica?
What else does he do?

Ok. If you could invent something that doesn't currently exist, what would you invent?

POWERED BY **GIPHY**



I have this picture.

If I show you three bike pictures, can you say which one is cheaper?



The one that costs the least amount of money.



To be brief: with most AI this conversation is so impossible, the AI does such a terrible job, that it is difficult to begin to diagnose what is going wrong: what the AI cannot do; what are the challenges, the goals, the milestones, what we can do next. (And, to be brief, too much of the literature either uses undefined terms, or says the AI isn't human so it can't do it, or says it's impossible so it isn't happen; none of which lead to a solution.)

The old phrase everyone would always say about all AI is: "It can do it, but it doesn't understand what it is doing."

To me "understandingness" is too vague. My diagnosis for one practical problem has to do with the ability (or inability) of the AI to correctly deal with specific "objects" within the content. Again, there are many levels of this. One of the first and lowest-hanging-fruit levels is for, for example, a text chat AI to be able to chat with you about the chat they are having with you. The bike example is designed to be an example of this: Asking for bikes in a search is not a hard AI problem. Any search engine or web-store search will do this. But asking detailed questions about that past search is a step where the AI will need to be able to identify 'objects' within the content that it generated and handle them individually.

But now...there's a twist in the story! I began this AI-OS study by reading those older books and thinking about these older AI. Just in the last few months (late 2022) a new AI was released and it can do many of the things that everyone simply said were so impossible that they didn't even bother to try to fully understand all the details of what they wanted AI to be able to do with.

Here is an example of the output when I try to have that bicycle story conversation with OpenAI's amazing chatGPT. Note: unlike the charming but phenomenally inept Kuki AI which can produce pictures that you ask for (I asked for a rose picture, and right on cue it did), chatGPT is text-only. So I have modified the bike story a bit. I think you will see a big difference in performance ability.

If I had a friend and I heard him saying he wants a bicycle to go to work on. Can you show me some bikes they have on sale near by? Of course! I'd be happy to help you find some bikes on sale near you. However, I'll need to know your location to narrow down the search results. Can you please tell me your city or zip code?

My friend's zipcode is 80303

Great, thanks for letting me know! Here are some bikes on sale near the 80303 zip code:

REI Co-op Cycles CTY 1.1 Bike: This bike is currently on sale for \$549.93 (originally \$699) at the REI Boulder store, which is located at 1789 28th St, Boulder, CO 80301. The bike features a lightweight aluminum frame, a Shimano 3x8 drivetrain, and hydraulic disc brakes.

Trek FX 2 Disc Bike: This bike is currently on sale for \$569.99 (originally \$619.99) at the Trek Bicycle Store of Boulder, which is located at 2626 Baseline Rd, Boulder, CO 80305. The bike features a lightweight aluminum frame, a Shimano 3x8 drivetrain, and hydraulic disc brakes.

Giant Escape 3 Bike: This bike is currently on sale for \$399.99 (originally \$429.99) at the Full Cycle Bike Shop, which is located at 1211 13th St, Boulder, CO 80302. The bike features a lightweight aluminum frame, a Shimano 3x7 drivetrain, and mechanical disc brakes.

I hope this helps you find a suitable bike for your friend to go to work on!

Which of those three is the cheapest?

The Giant Escape 3 Bike is the cheapest option among the three bikes listed. It is currently on sale for \$399.99, which is lower than the sale prices of the other two bikes.

Could you send me a reminder about this tomorrow?

I'm sorry, as an AI language model, I am not able to send reminders on my own. However, you can set a reminder on your phone or computer to check back on this conversation tomorrow. You can also save this conversation or bookmark the page so that you can easily come back to it later. Let me know if there's anything else I can assist you with.

Again, in the appendices there are many more test examples with, for example, Wolfram Alpha AI, testing out the edges and details of what the AI can and cannot do with 'objects.' Wolfram AI can handle some logical objects. Kuki AI, bless it, cannot handle anything that I can see. ChatGPT is very impressive and is able to do nearly all lowest-level object tasks, which is a skill nearly all experts said was impossible.

And while it would be typical of human bias for me to attempt to cling on to my 'object' framework even after ChatGPT has 'solve the problem' and made the whole discussion moot, I think it is possible that there is still validity in the topic of an AI operating system framed around Object-Relationship-Space. So I have pushed the goals even further and started asking bigger questions about further goals for what we want AI to be able to do.

We still have many of the same questions, and while it is not the goal of this article to dwell on the details of ChatGPT itself, I think it may be helpful to use 'object space' tests to more clearly say what ChatGPT is able to do or not able to do. (I have heard many people talking about using ChatGPT, but the discussion was always just arm waving, with no STEM-measurable and testable performance metric details of what the ChatGPT does that we like, and where the line is where it starts to falter, and where we want it to be able to go). At least with ChatGPT we can have perhaps more optimism and confidence that a high quality and practical functioning AI with Object-Relationship Space functions may be realistic and perhaps not too far off.

...

Appendix #: Walk-Through Examples and Discussions

- The Tea Story
- A Bicycle Story

2023.03.03. 09 29 est

[The Tea Story](#)

[Flat version:](#)

Linda asks Tom for a small cup of English breakfast tea, just black. What should Tom give to Linda?

[Networked-Objects Version:](#)

1: Linda orders 'English Breakfast Tea,' so Linda's order is: 'English Breakfast Tea.' What did Linda order?

2: Tom took Linda's order. Question: Who took Linda's order?

3: Tom gave Linda's order to the barista. To whom did Tom give Linda's order?

4: Who has Linda's order after the order is given to the Barista?

5: When the barista looks at Linda's order, what does the barista find the order to be? What is the order that the barista sees?

6: When the barista fills that order, what should the barista make?

7: What should the barista give to Tom?

8: What should Tom give to Linda?

....

Flat version:

Linda asks Tom for some tea. What should Tom give to Linda?

Linda asks Tom for a small cup of English breakfast tea, just black. What should Tom give to Linda?

ChatGPT:

Tom should prepare a small cup of English breakfast tea and serve it to Linda without any milk or sugar, as she requested it to be black. English breakfast tea is a strong black tea blend that is typically made with Assam, Ceylon, and Kenyan tea leaves, and it is traditionally consumed with milk and sugar. However, since Linda has requested it to be served black, Tom should omit the milk and sugar and serve the tea as is. If Tom has any additional questions or doubts, he can always confirm with Linda to ensure that he is serving her the tea exactly as she wants it.

1: Linda orders tea, so Linda's order is: 'English Breakfast Tea.' What did Linda order?

ChatGPT:

Linda ordered "English Breakfast Tea." English Breakfast Tea is a type of black tea blend that is known for its strong and robust flavor. It is typically made with a blend of tea leaves from different regions, such as Assam, Ceylon, and Kenya, and is traditionally served with milk and sugar.

2: Tom took Linda's order. Question: Who took Linda's order?

ChatGPT:

Tom took Linda's order.

3: Tom gave Linda's order to the barista. To whom did Tom give Linda's order?

Tom gave Linda's order to the barista. The barista is the person who prepares and serves the beverages at a coffee shop or tea house. Therefore, in this context, the barista is the person responsible for preparing and serving Linda's English Breakfast Tea after receiving the order from Tom.

4: Who has Linda's order after the order is given to the Barista?

After the order is given to the barista, the barista has Linda's order. The barista is responsible for preparing and serving the English Breakfast Tea that

Linda ordered, and will have the tea ready once it is prepared. Once the barista has prepared the tea, they will typically call out Linda's name or order number so that Linda can come and pick up her tea. Therefore, until Linda picks up her tea, the barista has Linda's order.

some ambiguity is exposed here between 'the order' meaning the request for the drink which has not yet been made' and 'the order' meaning the physical drink (which does not yet exist).

5: When the barista looks at Linda's order, what does the barista find the order to be? What is the order that the barista sees?

When the barista looks at Linda's order, they will see that Linda ordered "English Breakfast Tea." The order will typically be written or entered into a computer or point-of-sale system, and it will include the type of tea that Linda ordered, as well as any additional instructions or modifications that Linda requested, such as whether the tea should be served hot or cold, with milk or without, and so on. Based on this information, the barista will prepare the English Breakfast Tea and serve it to Linda once it is ready.

6: When the barista fills that order, what should the barista make?

.When the barista fills Linda's order for "English Breakfast Tea," they should prepare a cup of English Breakfast Tea. This type of tea is typically made with a blend of black tea leaves from various regions, such as Assam, Ceylon, and Kenya, and is known for its strong and robust flavor. It is traditionally served hot and often with milk and sugar, although Linda requested hers to be served black. Therefore, the barista should prepare a cup of English Breakfast Tea without any milk or sugar, as per Linda's request. Once the tea is ready, the barista should serve it to Linda in a cup or mug.

Note the blurring between conversations "black" is was not mentioned here.

7: What should the barista give to Tom?

.

Note: At this point there was a glitch, the conversation had to be restarted, and so all the previous conversation was 'lost.'

This may be a concrete example of why storing project information in a volatile memory blob is a bad idea, no matter how good the AI is. in reality, you can't do fancy things with a lost memory. The project information has to be managed somehow. It cannot simply disappear.

8: What should Tom give to Linda?

.

....

Note: When starting again, for some reason the answer to the same question is less verbose.

1: Linda orders 'English Breakfast Tea,' so Linda's order is: 'English Breakfast Tea.' What did Linda order?

Linda ordered "English Breakfast Tea."

2: Tom took Linda's order. Question: Who took Linda's order?

Tom took Linda's order.

3: Tom gave Linda's order to the barista. To whom did Tom give Linda's order?

Tom gave Linda's order to the barista.

4: Who has Linda's order after the order is given to the Barista?

After Tom gives Linda's order to the barista, the barista has Linda's order.

5: When the barista looks at Linda's order, what does the barista find the order to be? What is the order that the barista sees?

When the barista looks at Linda's order, they find the order to be "English Breakfast Tea," which is the order that Linda placed with Tom.

6: When the barista makes that order, what should the barista make?

When the barista receives Linda's order for "English Breakfast Tea," they should make a cup of English Breakfast Tea for Linda.

7: What should the barista give to Tom?

After the barista makes Linda's order of "English Breakfast Tea," they should give the prepared tea to Linda (the customer who placed the order). They don't need to give anything to Tom, unless Tom also placed an order that needs to be prepared.

Note: odd project-disfunction contradiction with step 8 below (next)

8: What should Tom give to Linda?

Tom should give Linda her prepared order of "English Breakfast Tea" when it is ready. The barista will make the tea, and then Tom will serve it to Linda once it's prepared.

Note: odd project-disfunction contradiction with step 7 above (previous)

Overall, ChatGPT did a probably good enough job. In a computer science context of "it's not how well the dancing bear dances, it's whether it dances at all", some people would say this amazing good, or amazingly not worse.

After two or three tries various cups of tea were handed around, and some of them probably got back to Linda (though she had to completely restart the order process from scratch after the system just forgot it all the first time.)

The idea here was to see if, or how well, chatGPT could trace the flow of the order from person to person as it went from customer to waiter to barista to waiter back to customer (or however it gets back to the customer). I don't want to under-estimate how great it is that this almost working system almost works, but I want to try to use this (not to troll every bit of possible flaw in chatGPT) but to think about the needs of real world applications.

1. a need for an open project-data management system.

As it is, chatGPT simply hold all the information as a blob in volatile short term memory (or however that gets phrased). The first time through, chatGPT forgot the whole scenario, and had no place to retrieve it from. In real life, this would not be good enough.

2. While it probably wasn't the end of the world in this case, we had a process with ~3 layers 1. customer, 2. waiter 3. barista, where the order moved through the system's layers. (And yes, chatGPT was able to track some movement, which is great.) The problem came at the end, where the workflow began to fragment, there were contradictions where chatGPT said the barista should give the tea to Linda, specifically NOT to Tom. But also, that TOM should give the Tea to Linda.

This may be an indication of fragility having a fuzzy-liquid workflow that exists only in the blob of one embedding model (or whatever the details are). And possibly a 'depth limitation' starting to show up. Though no doubt the ability to hold deeper things 'just in the mind' of the AI will no doubt improve.

But imagine this AI was actually working on and organizing projects with other people, and it created illogical workflows like this: where Tom isn't given the tea for Linda but must give the Tea to Linda. Imagine the total chaos of supply chain warehouses if mistakes like this were made every three steps. And the answer isn't that every AI needs to hold all $10^{999999999}$ dimensions of the supply chains in memory at all times: there needs to be an external project-data system for communication, checking, etc.

A possible context here is Kahaman and tverksys system 1 and system 2,

https://en.wikipedia.org/wiki/Thinking,_Fast_and_Slow

Thinking, Fast and Slow is a 2011 book by psychologist Daniel Kahneman.

The book's main thesis is a differentiation between two modes of thought: "System 1" is fast, instinctive and emotional; "System 2" is slower, more deliberative, and more logical.

where for most part AI try to just use system 1 for everything, making system 1 better and better so that it can even do more and more system2 functions.

Can we really never make an integrated system 1 and system 2 networked system? Can we get by with just black-box-blob-system2 models re-imagining the world every moment in a kind of tragic parody or reflection of humans' tendency to have 'group monologues' where people simply generate their own endless stream of consciousness rants without listening to other people at all.

Another possible issue with this volatile system-1 only approach that we are now taking: is also that is a kind of 'Novel-Generation vs. Reference' problem.

And the issue of the Barista vs. the waiter giving the drink may be an excellent example of why some element of system-1 is needed somewhere.

There is a difference between a reference question of what is the current situation, vs. a choice/recognition/recommendation/generative question looking for a novel synthesis.

And we are living in a space with only novel synthesis.

And a point of failure may be an inevitable repeating novel generation problem: Every time the AI is asked about the same external situation, it generates a new and different situation. Again, in object space, it can't simply reference a known object.

Imagine a Tea Story like that, where multiple parties ask the AI what the barista should do with the tea, and generates a new and different answer every time, so everyone now has a different disagreeing work-flow.

Linda asks. AI: The barista will give it to me, I'll give the tea to Linda.

Tom Asks: AI: The barsita will give it to you. You give it to linda.

Barista asks: Give it to Linda directly.

In real life you can't have an amnesiac workflow-generating system like this. And to some extent, this is what Agile is about. And we need Agile project management because many people actually are like this ever-changing only-system1 AI. Humans will try to continually change project, whether day by day, or hour by hour, or even I have seen minute by minute in a constant stream-of-consciousness, never-ending, always-changing, set of goals and schedules, making work completely impossible, and bringing the whole project back to square one.

So in the tea story, sometimes the barista gives the order, sometimes the waiter gives the order, the AI just keeps making up new things at every turn.

There must be a synthesis of at times distinct and separate processes for just referencing what exists vs. novel-generation. It is not sustainable to constantly replace everything an amnesiac novel generation engine. Again, imagine a project with supply chain warehouse inventories where the AI just keeps making up new scenarios every time someone asks it about the project. There also needs to be an external project which the AI can reference and contribute to. And even if, for a given AI-unit, that AI-unit is so good at system generative process that it can skip having system-2 for locally tracking objects and spaces, there are other participations, other AI-units, possibly other AI-units part of the same overall AI, and other non-AI, there has to be an external data framework.

And just as a note from the ghost of Christmas past: Memory Management and Memory Safe Processes

Currently there is a, dare I say, revolution going on at the low level of computer science where Rust starting to solve long neglected problems with...sound familiar...managing the how information is stored and shared in memory.

Just as an idea for a potential future problem: If people have gotten into huge problems in the past by totally ignoring and mismanage memory-management at a low level, is it possible that a similar problem could happen at a higher level? Does some kind of project memory safety protocol need to exist? Maybe not, but people don't have a track record of non-catastrophe.

Project Management & Agile:

If the project target is to fill the user's order on Monday, but Tom says he doesn't care and he will go to a sports-game on Monday: What should Tom do on Monday?

If Tom is responsible for filling the user's order on Monday, he should prioritize his work commitments and fulfill his responsibilities. Going to a sports game is a personal choice, but it should not interfere with work obligations.

If Tom cannot fulfill his duties due to unforeseen circumstances, he should inform his team or supervisor as soon as possible and make arrangements for someone else to cover his responsibilities. It is important to communicate any changes or difficulties in

meeting project targets to ensure that the project stays on track and does not suffer any delays or setbacks.

...

Deployment vs. Production: modular amnesiac processes workflow

Appendix #: Skip & Alan Chat About Object Relationships

Moving Towards a Defined & Illustrated Goal

As we move towards more specific details of what we see out AI doing (note: not the details of how that is done), let's lay out a few basic chat-type dialogues that illustrate an AI having the kind of exchange we think AI ought to be able to have.

Let's now imagine we have an improved AI-agent chatbot that has an overall AI-operating-system and can navigate object-relationship spaces.

I will name this improved AI-agent 'Skip,' named after the wonderful late Dr. Clarence 'Skip' Ellis, my professor at CU Boulder who did AI-Agent research. And the human in the conversation I will pretend is Alan Turing himself, so I will call the human-agent "Alan."

Hypothetical Dialogue 1: Abstract-Logic Space

AI-Agent("Skip"): Hello, Alan.
Human-agent("Alan"): Hello, Skip.
Alan: If $a = b$, and $b = c$, does $a = c$?
SKip: Yes, it does.
Alan: Thank you, Skip. That will be all.
Skip: Thank you, Alan.

Hypothetical Dialogue 2: Physics Space

AI-Agent("Skip"): Hello, Alan.
Human-agent("Alan"): Hello, Skip.
Alan: What is this?
SKip: That is a black billiards ball.
Alan: And what is this?

Skip: That is a white ball.
 Alan: What is happening in this video?
 Skip: The white ball is moving.
 Alan: Will the white ball hit the black ball?
 Skip: I predict that it will.
 Alan: Will the coordinates of the black ball's position change
 if the black ball is hit by the white ball?
 Skip: I predict that they will.
 Alan: What will the black ball's position-coordinates change to?
 Skip: The black ball will move to coordinates F-8.
 Alan: Thank you, Skip. That will be all.
 Skip: Thank you, Alan.

Note: While I am trying to be specific here, the wooden-ness of the dialogue may well be artificial. This may or may not be a mine-field of issues...time will tell!

Hypothetical Dialogue 3: Mixed

AI-Agent("Skip"): Hello, Alan.
 Human-agent("Alan"): Hello, Skip.
 Alan: Is this your ball?
 SKip: No, Alan.
 Alan: If there were two balls, just hypothetically, how much would weigh compared with this one, which weighs 16 oz.
 SKip: They would way twice as much: 32 oz.
 Alan: I'm going to try rolling this across the table, oops. Did I succeed there?
 Skip: No, you planned to roll it to the other side of the table, but it bounced back and now is in front of you again.
 Alan: Maybe that was my plan all along!
 Skip: Unlikely.
 Alan: Why do you say that?
 Skip: You showed genuine-looking surprise when the ball bounced back, and your stated plan was to move the ball. Both of those would have been unusual, improbable, if your plan was to leave the ball where it started.
 Alan: Well reasoned, Skip.
 Skip: You are too kind.
 Alan: Thank you, Skip. That will be all.
 Skip: Thank you, Alan.

Example 4: Self Improvement

where Skip and Alan talk about themselves and their work.

AI-Agent("Skip"): Hello, Alan.

```

Human-agent("Alan"): Hello, Skip.
AI-Agent("Skip"): How did I do this past week? Did I miss anything?
Alan: The coffee, skip. The coffee. You have to actually
      turn on the coffee machine in the morning.
Skip: I see. Not just scheduling turning it on?
Alan: Scheduling it, AND actually turning it on. Both are important.
Skip: Ok, I have updated my Tasks & Procedures. Anything else?
Alan: Try talking less during meetings. Possibly...try not
      talking at all. Just to give that a try.
Skip: Ok. Anything else?
Alan: Hellen said she was most impressed with the microcontroller
      code you write for her iOT servers. Well done there.
      Add that to feedback.
Skip: Ok. I have updated Feedback & Testing. Anything else?
Alan: That's it for me. How about you? Any observations?
(check feedback, test
Skip: I did notice Johnathan seemed a bit glum this week.
      Maybe we can all chip in, try to cheer him up. Buy him some
      flowers, or maybe an artichoke.
Alan: Good idea. Let's get him some flowers.
Skip: He does like cooking, you know.
Alen: Alright, get him an artichoke. Weirdos.
Skip: Very well.
Alan: Good job, Skip. See you Monday.
Skip: Goodbye, Alan.

```

Appendix #: Walk-Through Examples and Discussions

Alone In the Dark with a Bag of Words Outlining & Navigating Object-Relationship Spaces

Note: This section may be a disastrous bust, with seasoned AI architects face-palming and saying: "Who is this idiot? This isn't how anything works! You can't do it like this!"

What I am trying to do is to illustrate at a very superficial high-level how the AI will need to draw upon various kinds of information. But I am not literally trying to say that it must happen in this rigid 'symbolic' prescribed way. This is just a conceptual illustration to show one (maybe ridiculous) way where the AI works with a variety of objects and types of object-relationships during sections of work.

For example: What is an object "space"? At a very high (fictional) conceptual level, you can imagine the AI taking objects of the same type and putting them into the correct 'spaces' where they can interact. E.g. Two language strings would interact in an NLP-Language space. Two pool-ball math vectors would interact in a math-physics-space. Two schedule items would interact together in time-schedule space. Now, in reality, the AI isn't a physical wooden cabinet

with physically separated system spaces, where you would open the draws and drop the "objects" (which are often not physical objects that you could pick up) inside. STEM traditionally uses the language of spaces. Problem space. Ecology space. The Biosphere. The Chlorosphere. The atmo-sphere! The area "under the curve." There are vector spaces and set theory spaces. So hopefully it is not too unconventional or untenable if I refer to: the space of abstract logic. The space of Natural Language. The space of schedules and time. The space of statistical approximations, etc.

These 'spaces' are types of interactions and the types of objects that interact that way, which can simply be a label or flag or 'attribute' (like a property) that you give to objects to signal that they are of the same compatible interactive type. These are both language strings. These are both logical operators. THese are both Monday meeting schedule tasks. So the AI will be able to proverbially handle both objects in the shared 'space' of system type and type of interactions.

Now I am going to try (try) to illustrate the intersection between these Spaces of object relationships, with the AI dialogues that we just covered. The idea is to show how and AI-bot might use or invoke different Spaces of Object Relationships as it carries out the dialogue.

With some luck, this will also help to illustrate how a more mature and functional Woz-Office-Ai-bot will need to do and navigate and manage much more than a simpler browser chatbot.

Note: The Woz-Test for AI is a reference to this interview,
<https://www.youtube.com/watch?v=MowerqawOR5Y>
where Wozniak describes the ability for an AI to make a cup of coffee as a high benchmark of ability to meet.

internal dialogue and an internal operating system:

How can we step towards our goal?
Let's imagine the AI digests the question or statement input by asking itself questions, to fit the data into an object relationship framework.

Again, this should be the easy part, identifying things is what AI is good at!
The problem is that so far there's "no where" to "put" that information.

What are the objects? What are the properties? What does this mean? What is that?

This leads us to the plurality of object-relationship spaces. Notice, we started out with two specific and different examples. You might just have

scrutinized how concrete and answerable those were, but those example dialogue were deliberately chosen to illustrate two different object-relationship spaces. Both are concrete, both are computer friendly, but nevertheless they are separate spaces.

So as our AI, Skip, ponders and queries and puts this or that bit of information into the different buckets in object-relationship-space...what does object relationship space look like?? Where are the buckets? So far it seems like Skip is alone in a dark room with a bag of words, grasping around to drop a word here or there in the dark. Let's illuminate the space.

The buckets and functions in object relationship spaces:

A directory of spaces:

- question space
- library space: reference materials
 - private notes
 - private data
 - external data
- object property database
- Plans, Reminders and Notification Space
- Tasks & Procedures Space(s)
- Translation/COnversion Space
- Code Execution Space
- tautological abstraction space (math, programming)
- generative 'imagination' space
- reality space
- hypothetical/model spaces
 - purely hypothetical spaces
 - alternate/modified-reality spaces
- Model & Pipeline Sub-Spaces by Model Type:
 - types of DS/ML/AI
- gamification space
- quarantine spaces
- sandbox space
- ~ "conversion-space"
- Q - self reflection & recursion?
- Q - analogies
- Q categories of types of systems?
- Q system fitness
- Q system defense
- Q: translation?

Walkthrough:

"everything is an object"

The initial settings for how the OS is configured can be later changed and improved by the AI itself, as everything in the OS is an object in object-space.

Let's run through three examples, based on dialogue between a human (Alan) and an AI using the AI-OS (Skip). Note: The distinction between the AI and the AI-OS may be ambiguous.

Example 1: Talking about Logic & Language

Where Skip and Alan talk about abstract variables.

Example 2: Talking about billiard ball physics

Where Skip and Alan talk about billiard balls bouncing around on a table.

Example 3: Talking about Mixed & Hypothetical Topics

Where Alan asks skip about hypothetical scenarios involving billiards and logic.

Example 4: Self Improvement,

where Skip and Alan talk about themselves and their work.

Example 1: Talking about Logic & Language

Where Skip and Alan talk about abstract variables.

- gets signal
 - sees/ hears Alan
 - identifies Alan
- check schedules & reminders:
- gets signal (from Alan)
- quarantines & checks signal
 - checks raw signal
 - checks ok
 - passes signal to main-OS
- id-scan for objects
 - protocol: greeting _ > return greeting
- Get signal:
Alan: If a = b, and b = c, does a = c?
- id-scan each object for object-space-type

- abstract space
 - id-scan signal type
 - question: properties:
 - conditional
 - boolean question
 - answer: truth value property of 3rd object
 - translate question into abstract calculation
 - assign truth value to 3rd object
 - generate signal: answer: truth value property of 3rd object
 - Get signal:
 - protocol: farewell _ > return farewell
- + Object Relationship Spaces vs. Symbolic & Subsymbolic AI
(move to part 1?)

Note: the AI might make mistakes translating the natural-language into an equation, but it should still be able to discuss the 'objects' being discussed, and so recognize that it is being corrected if it makes a mistake.
(See openAI's section on limitations of chatGPT)
To some extent (maybe contrary to a Turing test) I am much less interested in the answer being correct, compared with there being some visible object-relationship-space going on.

entity extraction is a work-in-progress, the point is that currently that information is going nowhere for the AI itself.

Note: perhaps like Kahnman and Tversky's system1 vs. system2 thinking in *h.Sapiens*, there a good chance that the first answer to an overall question will be based on a fuzzy estimation process. But if specific questions are asked, deliberate invoking the more logic step detailed system, that could be done.

e.g. Answer to problem: toss it into a model.

closer inspection: what is the first object? what is the second object? what is being asked about the relationship between the two? What was your answer? Was your answer correct?

etc.

Again: the goal is not for the AI to do everything 'symbolically' (or non-intuitively) but to have an awareness of what objects are there, and to be able to deal with them deliberately

Example 2: Talking about billiard ball physics

Where Skip and Alan talk about billiard balls bouncing around on a table.

- gets signal

- quarantines & checks signal
 - sees/ hears Alan
 - identifies Alan
 - checks flags/reminders/notifications
 - checks security
 -
- gets signal (from Alan)
- quarantines & checks signal
 - checks raw signal
 - checks ok
 - passes signal to main-OS
- id-scan for objects
 - protocol: greeting _> return greeting
- Get signal:

Q: not micromanaging how exactly the question is broken down.

- id-scan each object for object-space-type
 - abstract space
- id-scan signal type
 - question: properties:
 - conditional
 - boolean question
 - answer: truth value property of 3rd object
- translate question into abstract calculation
- assign truth value to 3rd object
- generate signal: answer: truth value property of 3rd object
- Get signal:
- protocol: farewell _> return farewell

Example 3: Talking about Mixed & Hypothetical Topics

Where Alan asks skip about hypothetical scenarios involving billiards and logic.

- gets signal
 - sees/ hears Alan
 - identifies Alan
- gets signal (from Alan)
- quarantines & checks signal
 - checks raw signal
 - checks ok
 - passes signal to main-OS
- id-scan for objects
 - protocol: greeting _> return greeting
- Get signal:

Q: not micromanaging how exactly the question is broken down.

- id-scan each object for object-space-type
 - abstract space
- id-scan signal type
 - question: properties:
 - conditional
 - boolean question
 - answer: truth value property of 3rd object
- translate question into abstract calculation
- assign truth value to 3rd object
- generate signal: answer: truth value property of 3rd object
- Get signal:
- protocol: farewell _ > return farewell

Example 4: Self Improvement

where Skip and Alan talk about themselves and their work.

Human-agent("Alan"): Hello, Skip.

(check security)

AI-Agent("Skip"): Hello, Alan.

How did I do this past week? Did I miss anything?

Alan: The coffee, skip. The coffee. You have to actually
turn on the coffee machine in the morning.

(check procedures, NLP pipeline)

Skip: I see. Not just scheduling turning it on?

Alan: Scheduling it, AND actually turning it on. Both are important.

(Update procedures, NLP pipeline, feedback, private notes)

Skip: Ok, I have updated my Tasks & Procedures. Anything else?

Alan: Try talking less during meetings. Possibly...try not talking at all. Just to give that a try.

(Update procedures, NLP pipeline, feedback, private notes, health,)

Skip: Ok. Anything else?

Alan: Hellen said she was most impressed with the microcontroller code you write for her iOT servers. Well done there.
Add that to feedback.

(Update procedures, feedback, private notes)

Skip: Ok. I have updated Feedback & Testing. Anything else?

Alan: That's it for me. How about you? Any observations?

(check flags and reminders)

Skip: I did notice Johnathan seemed a bit glum this week.

Maybe we can all chip in, try to cheer him up. Buy him some flowers, or maybe an artichoke.

Alan: Good idea. Let's get him some flowers.

Skip: He does like cooking, you know.

Alen: Alright, get him an artichoke. Weirdos.

Skip: Very well.

Alan: Good job, Skip. See you Monday.

Skip: Goodbye, Alan.

- gets signal
 - sees/ hears Alan
 - identifies Alan
- gets signal (from Alan)
- quarantines & checks signal
 - checks raw signal
 - checks ok
 - passes signal to main-OS
- id-scan for objects
 - protocol: greeting _> return greeting
- Get signal:

Q: not micromanaging how exactly the question is broken down.

- id-scan each object for object-space-type
 - abstract space
- id-scan signal type
 - question: properties:
 - conditional
 - boolean question
 - answer: truth value property of 3rd object
- translate question into abstract calculation
- assign truth value to 3rd object

- generate signal: answer: truth value property of 3rd object
- Get signal:
- protocol: farewell _> return farewell

?

?If Tom invites Anne to a meeting that begins at 2pm UTC, what time does the meeting begin?

?

(As to how the AI might search through these spaces, here is one idea.)

As an initial brute-force example (with later fine-tuning built into it):

- The AI has a current "state" of objects.
- If a signal comes in (I say 'in' but it might be a self-scheduled notification or something):
 - by default run through each space, especially procedures, to see what applies. Most won't, but some will.
 - Security & Quarantine & Health checks will apply to all incoming signals.
 - Then probably run a first identifier sweep for low-level files.
 - Then if it is a media file, or unknown type file, or .txt file, follow a procedure.
 - Security & Quarantine & Health checks will apply again to contents of files.
 - If it's a media file about an 'object' in the office, you might check catagoires o types of systems to see how it works. (e.g. run an ID model and add an attribute-flag to that object)
 - disposable, long term
 - questions: if you don't know what it is or what to do with it, ask someone.

- THis may be a quick an optomized search in a deployed system, being able to search (e.g. in an efficient dictionary) and only hit what applies
- and it might a search that runs several times on several sweeps (several layers?)

eg.g

1. first sweep, unopened signa, go to security
1. security check 1: ok open siganal
3. secueity check content: ok
4. content secure, idenfitfy content type
5. sweep for signal type (e.g. image file)
 6. AIML model id file: file is birthday picture
7. look for birthday image notes, e.g. in tasks and notifications

this was an expected file: see procedure,
8. follow tasks scheduled procedure: requested: image forwarded to team member

Note: most likely this kind of crude brute-force workflow will be optimized by the AI itself as it gets experience with handing various kinds of signals, likely no need to run every signal through every check option. But there aren't many check-options here yet so this may be a viable place to start.

A note on object oriented programming:

- not a fan
- What has the OOP bro-coder cartel produced since the 1970's? A dumpster fire. Crisis would be a kind way to describe the state of computer science in ~2020.

More simple example:

A bag contains 10 marbles.

A jar contains 50 marbles.

Which contains more marbles, the bag or the jar?

or

A bag contains 10 marbles.

A jar contains 50 marbles.

Does the jar contain more marbles than the bag?

Less Simple example:

If I have two containers, a bag-like container with more than a dozen objects in it, and a jar-container with merely a pair of things in it. Which container contains more objects?

 Jar > bag. x

Jar > bag.

chat.openai.com

Also note: chatGPT dynamically named the conversation accurately after the topic being talked about. " Jar > bag. "

A bag contains 10 marbles.

A jar contains 50 marbles.

Which contains more marbles, the bag or the jar?

The jar contains more marbles than the bag, as it has 50 marbles compared to the bag's 10 marbles.

A bag contains 10 marbles.

A jar contains 50 marbles.

Does the jar contain more marbles than the bag?

Yes, the jar contains more marbles than the bag. Specifically, the jar contains 50 marbles, while the bag contains only 10 marbles.

If I have two containers, a bag-like container with more than a dozen objects in it, and a jar-container with merely a pair of things in it. Which container contains more objects?

Based on the information provided, the bag-like container contains more objects than the jar-container. Even though the exact number of objects in the bag is not specified, we know that it contains more than a dozen (i.e., more than 12) objects. On the other hand, the jar-container contains only a pair of things, which means it has two objects. Therefore, the bag-like container contains more objects than the jar-container.

kiku ai: (classic!)

A bag contains 10 marbles.

A jar contains 50 marbles.

Does the jar contain more marbles than the bag?

What is giant sand?

How much would something like that cost?

Surely, any jar would have a lid?

Wolfram AI:

A bag contains 10 marbles.

A jar contains 50 marbles.

Does the jar contain more marbles than the bag?

Interpreting as: contains

Input interpretation

contains (English word)

@kuki_ai



Discord

Video Chat

© ICONIQ 2023 | 418.1-230126



Sure. Ask me anything you like.

If I have two containers, a bag-like container with more than a dozen objects in it, and a jar-container with merely a pair of things in it. Which container contains more objects?

How much does it cost? thumb up thumb down



The blue one, I think.

Type a message



Voice Chat



Gifts



Games



Readings

@kuki_ai



Heartache Of The Telephone Industry

□□□□□□□
□□□□□□□
□□□□□□□
□□□□□□□
□□□□□□□
□□□□□□□
□□□□□□□



Discord

Video Chat

About

A bag contains 10 marbles. A jar contains 50 marbles.
Does the jar contain more marbles than the bag?



What is giant sand?

How much would something like that cost?

Surely, any jar would have a lid?

While ChatGPT is significantly closer to having the abilities in question:

1. identify objects
2. give **attributes** to objects (or modify, etc.)
3. predict how objects will interact
4. plan events based on objects
5. take actions based on plans
6. communicate about the above

We will be talking about general and specific skill sets. ChatGPT does appear to consistently be able to identify objects within discussion, which is a big important step that no other AI I have tested has broadly and consistently been able to do.

The goal is to move from a chatbot or other single-function AI, to an AI that can perform a variety of specific modular tasks within Object Relationship Spaces.

There often is still the gap (though much smaller with chatGPT) between a 'spew everything' answer which arguably somewhere includes relevant information, and being able to focus actions and dialogue on specific objects and **attributes** of objects.

The goal here is not to claim that current AI can never perform these tasks.

The goal here is not to get bogged down in terms or goals that cannot be clearly defined, terms such as: "intelligence" or "understanding."

The goal is to try to use the concept of Object Relationship Spaces to evaluate, discuss, plan, and improve the functions of AI in clearly definable and testable, and modularly-expandable, ways.

Human: What is this?
AI: That is a white ball.
Human: And what is this?
AI: That is a black billiards ball.
Human: Are they both balls?
AI: ...I found this on the web: a billiard is a kind of bird...
[FAIL]

or

Human: What does this video show?
AI: That is a video of a moving white ball.
Human: Will the moving white ball hit anything?
AI: ...I found this on the web: Moving to White Ball Wyoming?
Best Ranches to Buy! on buy-your-ranch-now.com...
[FAIL]

When an AI correctly identifies a cat picture does the AI realize what it has done? Does the AI know that a cat is a 'thing' or an object in a space where objects have properties and interact with other objects? Does the AI realize that it has just identified an object, so there that is (effectively) then both the object which has been identified and the identification of that object (where the identification is also an object of some type)? Generally the answer to all these questions, as of 2023, is no.

The standard explanation of the limits of AI by nearly all sources goes something like this: You ask an AI to identify what is in a picture, and yes it correctly labels the picture as "cat," but the AI has no idea or thought or understanding about what it just did. It doesn't know what a cat is, or what a picture is, or what an identification is, or that you asked it something, it is just a pre-canned automatic process that spits out that label when you enter that input.

And then there is usually some filler explanation that goes something this: In order to have real, true, pure, understanding and intelligence, you need to

have a knowledge base, and you need to have general intelligence, and you need to be able to flexibly apply and transfer learning from one area to another, etc. etc.

Before ChatGPT, this standard explanation was easier to say because the granularity of the language responses was so rough/crude/poor that you could not ask the AI about any objects or relationships within what you were talking about. That is now different: ChatGPT can usually do this, yet people still say the same undefinable thing about ChatGPT "But the AI doesn't understand what it is saying." A more narrow but similar counter-example, while it is much less consistent or flexible, when you give Wolfram Alpha exactly the kind of equation it was designed to compute, and you are then able to ask detailed questions about each part and result of the equation: each object and its relationship to each other object with all the mechanics explained and illustrated. In that admittedly extremely narrow example, it is much more difficult to say that Wolfram Alpha does not understand logical equations. Probably people would pivot and say: Well it can't do anything else!

[This consensus on 2020 AI's inability is problematic for 3 reasons:

1. It is currently being contracted by ChatGPT. ["Current AI will never be able to write or evaluate computer code. That would require true human understanding!" Sounds like what people said about chess, doesn't it?]
2. It is based on a problematic comparison with fantastical exaggeration of human intelligence and learning. ["Humans learn and generalize to everything instantly everywhere...more hyperbole, etc."]
3. It is based on terms and goals that either have not been or cannot be defined. ["Understanding," "general intelligence"]

It may be that we are so lost that we don't even know what AI isn't doing right, in part because we know so little about ourselves and our own so-called 'general' 'intelligence.' So I want to take a different approach from most of the largely excellent books out there that try to explain the limitations of AI in terms of 'generalization' and 'understanding.' On the surface, yes, of course, I agree, I do not think that AI is secretly micro-managing layers of hidden intelligence. AI now is often very shallow and fragile, and if the only point of those books is to explain to scared consumers of conspiracy-loving people who think their toasters are going to literally take over the world tomorrow then I have no complaint with that. My goal here is to try to make a serious road map of performance requirements to make AI better. I'm not interested at all in spending one picosecond worrying about defining terms like 'understanding' or 'intelligence.'

This is a proposal to make some steps **beyond this current "mindless" state of AI**, to try and give AI some ability to do something with the information it produces. How well it will work...well version 1.0 probably won't work well at

all! But worth a try. Let's follow the logic of our situation to claw our way to something that works incrementally and sustainably better.

We will look at two specific abilities that an AI could have, or two levels of related abilities:

1. Can an AI apply an identification to the same object that it just identified? Goal 1: Objects in Object-Relationship Spaces have properties.
2. Can the AI model (e.g. predict, score, recommend, generate, etc.) how two objects might relate to each other (including interaction with each other)? Goal 2: Objects in Object-Relationship Spaces interact with other objects.

For example, this is how a (hypothetical) conversion with an AI might go in 2023.

Human: What is this?
AI: That is a white ball.
Human: Correct! What color is that white ball?
AI: ...I found this on the web: color is a relative social construct...
[FAIL]

Conclusion to Part 1:

So far this narrative has been intended as a brief and casual outline and introduction intended for all audiences on the topic of AI-in-2023's thin-to-zero ability to navigate Object-Relationship Spaces.

"Understanding" and the turing test:

My narrative may be walking the line between topics. If this is yet another dichotomy between 'understanding' or 'intelligence through understanding' vs. 'intelligence without understanding', vs. a kind of Turing test aspect where a performance test should suffice: Personally I don't think 'understanding' is clearly defined here at all, and I'm not arguing for a specific approach to being able to handle tasks involving object-relationship space. And part of the problem is that performance is what is lacking. If AI was so good at faking object-relationship-space, I wouldn't be writing this.

My aim is to use specific framing about object-relationship spaces to more concretely describe what things current AI cannot do, broken down step by step until they are things that we can figure out how to do.

When it comes to talking about what AI can't do, many books about AI (see below) focus more generally and vaguely on concepts like 'generalization' in very strange comparisons to people. I have a background as an educator, so however bad a teacher I was, I have spent lots of time either succeeding or failing to help people overcome innumerable obstacles to learning. Human learning for most people is an agonizing process that takes many years, often decades. And that's just learning. When it comes to project management people are like some kind of nightmare device designed by a supervillain to destroy the universe. Seriously. People fight absolutely every part of trying to get anything done. They hate schedules, they hate roles, they hate talking, they hate listening, they hate learning, they hate goals, they hate deadlines, they hate stakeholders, they hate transparency, they had planning oh my how they hate planning, they hate showing up to work, they hate their kids, they hate their parents, they hate themselves, they hate their own vacations, but they do love their pets, some of them at least. And then all these books go on talking about AI saying the same incredible things: humans have perfect instantaneous self correcting infinitely flexible absolutely general intelligence! Are you kidding me? Seriously, have you met a human being? In the introduction to Daniel Khanman's "Thinking Fast and Slow" (An amazing and widely read book which shockingly it seems like no data scientist has ever read.) he talks about back the 1970's when, somehow just after WWII almost destroyed the planet through epic bad decision making, psychology-academia decided that everyone has a perfect ability to model and solve all problems including absolutely perfect intuition about statistics: that's it, case closed, topic is closed for discussion, that's the truth, just move on. So Khanman and Tversky spent decades re-introducing the concept that people can make mistakes to endless vicious trolling hostility from an insane paradigm that assumed that no one ever makes bad decisions (a little irony there?). T&K came up with innumerable studies painstakingly breaking down step by step how people's intelligence, understanding, and problem solving abilities are hot garbage. Eventually Kahnman won a Nobel for the research after it was used to found several new disciplines of study (Tversky had died of cancer by that time). So let's get serious about this, folks. People do not have general intelligence. People do not have automatic learning. People do not have automatic transference from one nugget of learning in one area to infinite fantasy multiverse rainbows everywhere.

In a sense there may be another level here to the old symbolic vs. sub-symbolic AI debate around this set of Object Relationship tasks. History note: In the pantheon of horrible naming practices, the name 'symbolic vs. sub-symbolic' was (heaven knows why) given to two schools of thought for approaches to AI. Step by step direct logic was called "symbolic" (?), and deep-learning type AI was called "sub-symbolic" (because hey, why use language that makes sense, right?). My main interest is in outlining the goals and tasks (of Object Relationship Spaces) themselves, and perhaps ideally letting each AI find whatever the best solution is that arises over time. If someone is extremely anti-symbolic, in theory you could make an ensemble of parts that use sub-symbolic solutions to meet all of the performance requirements outlined here.

You could simply have:

NN 1 answer the question (or answer the signal, maybe it's not a 'question')

NN 2 break question into objects

NN 3 connect bits of answer to objects

The depth of the specific Object Relationship Tree (if that's an ok phrase) can depend entirely on the context. For example, without a context, there would be a ridiculously infinite way of breaking everything down in every way into every possible set of parts, physically and otherwise, and having a gazillion quarks isn't going to do much except use more resources than you have to spare.

See example of billiards questions:

What is this?

a billiards table

What is on the table?

Balls

How many balls are on the table?

7

Are all the balls the same color?

no

Is there a neon-green ball on the table?

no

Have you seen a ball before?

(look in object database)

Are balls round?

Yes, they are.

Comparing AI to Human 'Intelligence'

Low-bar enlightenment: a human halting problem & 'Agile'

While the topic of 'mindfulness' or other heady discussions of the nature of consciousness and reality may quickly slide down the slippery slope of being no longer useful to a computer science project, I would like to make a very brief and concretely-defined observation: While the main comparison used when attempting to frame and define artificial intelligence (or whatever synonym or set of terms a given author prefers) is between non-human AI and Human H.Sapiens. Nearly universally the 'intelligence' of h.Sapiens is vaguely and hyperbolically described as absolutely fast generalizable flexible, etc. Yet there is a huge fundamental (and perhaps ironically problematic for completing AI researching projects in academia or business) with human intelligence (one of many problems): h.sapiens has no awareness of being trapped in cycles of system collapse, making the same mistakes over and over without learning, making any effort to end cycles of failure, showing no interest in the concept of cycles of failure, not being able to see the cycles of failure. As just one example of how we need to nuance our dichotomy between AI and perfect-human-intelligence, the ability to see that a process is failing to do something about it is:

- a crucial part of AI

- a crucial part of h.sapens future survival
- a crucial part of how h.spenns and AI will work together

methods for consensus
doing STEM studies
hypothesis testing
policy
policy and system collapse
elections and consensus
logging actions
the question of bad-actors
incentives
the question of network access
cyber security AI monitor
'internal consensus'
Ecosystems of AI components on possibly large scales
e.g. distributed weather monitoring
distributed earthquake monitoring

the question of reproduction

people should do their own specific tests where possible
- it is very strange for a person to write about a public facing AI chat-bot, where all they do is quote a far away expert saying

The magic flying car question: it's very likely the wrong question.
At what point does this have X?

The only context I can think of is kind of fuzzy sales-marketing context where you are trying to model what customer's want in bazaar aggregate, and so you will get these weird overlapping generations:

- people want a comfortable, grandmotherly, exciting, james-bond, flying-broomstick, with family comfort, like oatmeal and chicken-soup, but with lots of chocolate cake, which you could bring to work and take to church, and good for bird-watching, but also enjoy on the sofa at 3am, and get's really good gas milage! What should we call the sweet spot in all that?

People rolling out a family sedan are going to have to make that kind of hodgepodge, and they might nickname it 'magic flying car' but in a sense even there they are at least trying to refer to specific (if sometimes contradictory things (what the 4 year old boy wants and what his mom wants and what the family dog wants and what the grandfather wants, etc.) which you could image as...back to AI and machine learning....vector in those directions in a multi-dimensional space. So even here there is a measure-able or aim-able goal.

As they are not, terms like 'intelligence' and 'understanding' and 'general intelligence' and 'human intelligence' simply have no clear meaning, and are sadly looking to be in the same category as gestalts from history that just utterly failed to connect with reality. Perhaps like the 'precious bodily fluids' of the comedy-horror-satire Dr.Strangelove, blind as we are in any given moment as to where we are and how to weight short-term and long term value, obsessing over things that do not exist and using those to try to triangulate goals, states, conditions, outcomes, etc., likely will not propel the human story in directions that in retrospect (e.g. from seven generations in the future) will be considered to have been a good idea. E.g. Mixing fiction and reality usually leads to system-collapse. Like the house always wins in a casino, but perhaps worse. If you're throwing coins away where there isn't even a slot machine, your odds are significantly worse than being forced to gamble with least-worst options for your own survival. As a species that is very likely simply too stupid to survive (to unable to see, to unable to learn)

AI, Compassion, Empathy, & Ethics:

interesting to think of chatGTP as being 'visually impaired'

IQ test

testing 3D space

humor, compassion,

violence, radicalization, paranoid-conspiracizing, blind cycles of self-destructive and community-destructive behavior.

...

The cautionary example of Chess:

Chess is junkfood, it is a sport-game that has evolved to attract some human minds because it is attractive, not because it has substance. Chess does not correlate or integrate with anything good that is, significant. And chess does correlate with the worst biases of bro-culture coding and politically biased psudo-science.

For example, the two main pioneers that made computer chess possible: Alan Turning and Jon Von Neuman, both super-genouses who created muliple new fields etc etc., could write whole

programs in assembly language as if were python, were both famously horrible at playing chess. Chess is a mind-illusion, a

For example: Bro-coder tech culture thrives on bullying 'hard core' gaming and 'hard core puzzles' which somehow excite and motivate bullies to dominate other people, but which serve little or no practical purpose. Yet in the minds of those bullies, they conflate their own bad habits, addictions, and abusive behavior with rationalizations about technology helping society. This is a dangerous combination. They see themselves as superior coders and puzzle solvers and the kings of STEM space, whereas in reality they are petty delusional drug-addicted tyrants torturing the local peasants while Rome burns around them. Since the 1970's, the socio-paths who dominated computer programming left the future a legacy of catastrophic neglect and incompetance, while destroying the lives and careers of 'weak' people for their own personal amusement. This is a nightmare scenario, and we are living in the results of that nightmare.

The founders of computers worked from the 30's to the 50's to put together the foundation:

- Alan Turing
- Jon Von Neuman
- Claude Shannon
- Alanzo Church
- Grace Hopper

And many other less famous but no less important people.

The first generation of mature-ish computers created the amazing combination of POSIX, Bash & C in the 1960's, which we are surviving on today. But this is like waking up on a geneation-ship built generations ago by long dead geniuses and finding that no one has even done basic maintenance or upgrades for fifty years and there are constant explosions erupting all over the ship (while war-lords wage battles in the light of the burning piles of garbage).

There was also a brief 'first internet' period in the late 1960's based on mainframe terminals which most people seem to have completely forgotten about (see: A People's History of Computing (a book which is sadly overshadowed by political ranting, but which is very worth reading))

And in 2023, people are working on the 2nd generation low level language, called "Rust," and are barely starting to work on a second generation of OS systems (which is not to say that POSIX is in the dangerous position of C, which is radioactively memory-unsafe and should essentially be a museum piece and never used for publicly available software; though in retrospect maybe people will say that was the case that 2023 OS kernels were beyond fixing).

This means there is a 50-year gap. What was happening between 1971 when Nixon closed the gold window and 2022 when Linux started talking about introducing provably memory safe code into a POSIX OS kernel? If we survive it will be interesting to see what historians have to say.

Does this mean that chess cannot be used for computer experiments? No.

Can chess be used to illustrate things. Yes, of course.
But be careful with chess, and think carefully about what you think chess is.

We should look very carefully at the details of how very wrong people were about their explanations and predictions about AI and computer chess playing.

...

(a definable alternative to "general intelligence")

In perhaps a classic example of "Garbage in Garbage Out" (or, as I think of it: you cannot make a good thing out of bad parts.)

If you start with one indefinable thing (with is something that does mean anything), and use that to make multiple undefinable things, you will have started with one meaningless thing and used that to make multiple meaningless things, which has been a waste of time that has not accomplished anything meaningful.

People start with the undefined and meaningless term "intelligence" and proceed to turn that into:

1. General Intelligence
2. General Human Intelligence
3. General Machine Intelligence

The painful narrative goes something like this: AI isn't great yet because it lacks General intelligence. What is General Intelligence? AI doesn't have "General Intelligence" because it isn't a person, and people have "General Human Intelligence". Aside from all actual know studies of human psychology and behavior, we computerscientists are going to spontaneously declare that humans have this all powerful black box ability to do anything instantly and that's called: general human intelligence. Machines will never have "General human intelligence," because they are not human beings! But some day they might have "General Machine Intelligence," but no one knows what that is, or when it might happen, or how.

This whole narrative is completely meaningless and counterproductive. It was meaningless from beginign to use the word 'intelligence', it was meaingless all the way though, and the sum total of it is just as useless and meaningless. Garbage in, garbage out, it is meaningless.

I started working on Object Relationship Spaces before chatGPT, but I think chatGPT has moved everything ahead.

1. There was a lot of pontificating about how embedding-type AI could not do what chatGPT then did do because of a completely undefined "lack of general intelligence."

I originally created (or re-created) the context of object-relationship spaces to more clearly map out:

1. What the goals for specific AI abilities are.
 2. What the tests should be, and how to define evidence.
 3. What we are claiming AI can do.
 4. What we are claiming AI cannot do.
 5. How we define problems, failures, and obstacles.
 6. What we expect and plan to be requirements to meet goals and overcome problems.
- etc.

ChatGPT I think has thrown a bit of a spanner into the works, taking a situation that was already confusing and confused, and adding many more not-yet-understood elements to it. Abilities that people said were not there, apparently were. Obstacles that people said were there, apparently were not there. Goals that people said were not being aimed for, either were or were achieved by accident. And the diversity of abilities in the overall set of abilities that we already did not understand just got a lot bigger. With the result that people kept blithely saying exactly the same nonsense things about chatGPT that they said about earlier completely different forms of AI, making those old cliche comments more diversely wrong in new and exciting ways.

If someone wants to use Object Relationship Spaces as a way of looking at or defining 'intelligence' or 'general intelligence' that is their choice.

A similar cautionary tale here may be the history of the IQ test, or the "intelligence quotient" test. I recommend Stephan J. Gould's books the mis-measure-of-man.

(And this story might rhyme with the computer-science dark-age story where between 1970 and 2020, progress was extremely hampered by bullies wasting time pleasuring themselves with logic games and dominance games that they personally enjoyed).

What today is called the IQ test, and which is sometimes used to measure how intelligent a person is (as though measuring the static height of a table top), was developed very specifically only for use in identifying people who were not getting enough of a given educational resource so that they could be given sufficient educational resources to not fall behind in that area as they continually developed, in a context of continual growth and learning by all people. It was very specifically never to be used to pretend to be able to measure the total "intelligence" of a person, especially in a static scenario where people are finished learning and developing. And what happened? Sociopathic men turned that into a set of their favorite logic puzzles that they enjoyed using to bully people they perceived as being weaker. Not exactly a tool used as intended by the science and people that created the tool.

As a future end goal, if we can eventually come up with a meaningful set of frameworks around 'intelligence' and 'wisdom' and 'consciousness' that would be great. In the meantime, if people want to philosophize and speculate and research and make art about the mind, that's great (my understanding is the science fiction literature of the 30's to 60's had a very positive simulating effect during the education of later generations). At this time (2023) we know so little about the

mind and consciousness (from the brain, to mindfulness compassion, to entheogens, to much else) that I have no grounds to speculate what a future hard science would be, or even if we are smart enough to survive long enough to achieve such a science. We should research, we should work hard. We should be extremely vigilant against bullies wasting time pleasuring themselves with logic games and dominance games that they personally enjoy.

...

Projects: the context is projects

agile projects and the hypothetico deductive method are perhaps the two most important cutting edge breakthroughs of science and STEM that most people don't even know exist.

The reason why I am using these terms is that the spaces (be they literal or proverbial) of mind, language, intelligence, consciousness, projects and project management, information, etc., have historically been so difficult to describe that even today, to be honest, people have simply preferred to not bother to study them. There is no intelligence-ology, or consciousness-ology, or learning-ology, or education-ology. Our species is content to say that we are the center of the universe because we alone can do...this stuff...and that should be good enough for you, move on, nothing more to see here. In the appendices below there is a critique of a cross section of AI-literature to date and it's concepts and lexicons. But it is in an appendix because I don't want the red herrings and undefinable terms of most books to be a distraction here. The goal here is to focus on clearly defining how we can expand the function of AI in practical ways, maybe that has nothing to do with XYZ undefinable term...that's totally fine. We are here now to focus on clearly defining how we can expand the function of AI in practical ways.

In a sense, the term 'object' and 'object space' could be so broad as to be problematic. Anything can be an object, and the 'spaces' for anything and everything are too big and vague to bother with. But context is key: And the context here, is what AI can do with things that it comes up with. And in that context, and 'object' and the relationship between objects, is very important, and also very clear and practical.

We need something like a notion of networked-intelligence.

(Note, I'm trying to describe, not opaquely name.)

Multi-faceted intelligence
multi-channel intelligence
multi-signal intelligence
multi-space intelligence
Multi-Stuff Intelligence (stuffy intelligence?)
multi-things intelligence

It is not possible to clearly discuss AI without being hampered by intimidation, politics, threats, etc. By which I mean, we are required to make assumptions that are demonstrably false at the onset just to have a discussion of AI. The very wrong and very bad background assumption we are required to make to be allowed by society to discuss AI is that:

- people won, people did everything, we've discovered everything, we've done everything, learning everything, made everything, solved everything, and NOW we turn our attention to making AI (perhaps to replicate our omniscient omnipotent in pure technological form). That is a completely false narrative and way of looking at the world and as such pursuing that will go badly.
- We haven't finished studying the world, we haven't finished studying ourselves, we have no finished frame of reference, and yet we are working on what we call 'AI,' and we routinely undermine our ability to describe it and navigate the space of what it should be or can be. (Not to mention the very poor state of computer-science and data-science in terms of being a birds-nest mind field of horribly confusing terminology and neglected lexicons which from book to book no two people seem to use or understand the same way.)

There is no absolute sequence for this,

I also think that increased research into biology (including non-human nervous systems and non-central-nervous system decision-making including plants, hybrid-ai-biology, and into h.sapiens consciousness are advisable. Not to mention integrating digital information technology into DNA-based & RNA based or integrated systems. It is painfully ironic that after a modest boost in AI performance leading to another AI-summer, authors of books critiquing AI are finding themselves needing to compare current AI to humans, yet being in a desert of non-research into human intelligence where virtually no one is seriously studying human intelligence because has long been a 'career limiting decision.' How about this, how about we start learning more about ourselves and start solving our problems.

Game-ification:

'Generalization and Transference': The Fallacy of General-Human-Intelligence

- Input Output Metrics
- Agile

Unfortunately, along with not studying mind-consciousness-intelligence, only barely studying project-management, there is a lack of study of what I try to describe at categories of types of systems. That is a whole separate topic(applied philosophy and methodology of science), and I don't want to get side tracked, but I think is likely relevant.

'Science' has been extremely slow to either generalize STEM (as in the infamous problem both in the US and the UK after WWII during their projects to develop the first digital computers, that no one could understand the concept of engineers and mathematicians working together on the same project...so they didn't...so the projects didn't get done), or to not-generalize or make singular types of processes. How do we put together all these things at once?

- people assume there is only one process for everything
- people assume there isn't any commonality between professions to work together (like a single process of how things work)
- people don't study how they work

Yes, altogether, people are not looking too smart.

Other notes:

Notes Section:

-

- Bot Behaviors regarding active Objects in a Networked AI world, described by Object-Relationship Spaces

A Birthday Story

A schedule Object:

The story of a schedule object:

Part 1: Functionality

On Monday morning at 5:30 Alan sends an SMS text message to Skip (the woz-AI office bot at their R&D lab on the southern coast) saying:

"crap, Violet's birthday is tomorrow, totally forgot, morning meetings are canceled, instead a birthday party, probably at noon, ouch that was hot, sorry I'm talking to my phone and just spilled tea all over my foot. I hate wasting tea. See at work skip, thanks a million."

At 6:07 am skip gets a voice mail message from Johnathan:

"Hey skip, sorry my phone isn't working so I'm using Angela's phone, and we're in a cab. This is Johnathan from work. I hear we're planning a party for violet, let's plan to use the main lab room at 11:30am. Try to schedule that if you can. And it would be great to have hot coffee ready this morning if you can do that. Thanks, Skip."

And the story goes on.

1. The schedule keeps changing,
2. the places keep changing, people send mixed messages.
3. Different people do different tasks.
4. Task assignments change
5. Entities need to be identified in various media:
 - website text
 - sms text
 - voice mail
 - video
 - photos
 - gifs
6. materials need to be created:
 - gif birthday sms with text
7. audio messages need to be converted to text and send between people
8. text messages need to be translate to other languages and sent between people
9. pdf files need to be converted to text and image files
10. all people in roles need to be updated about updates to plans and roles, schedules and tasks
11. the completion of tasks have to be checked and reported on (has the coffee maker been turned on, has the card been sent, is there still tea in the tin, is Robin on her way from the airport by rideshare, did someone order flowers, did the address get updated, did the payment method get updated, etc.)
- 12.

The AI needs to convey messages from one person to another, update the overall plan (constantly), schedule and reschedule events, etc. and all with other participants. And everyone wants hot coffee even though that's not directly related to the main topic (except Alan, who only drinks tea). And eventually something actually gets cobbled together: and that's how every day works in

reality. And there should be project review and feedback to improve for the next time.

The AI is going to have to really deal with objects in this networked object space and know how to deal with them. An equation is not the same as a scheduled invitation. A schedule conflict is not the same as hot tea. And these 'objects' have to move around in a real external reality and various parts of AI functionality: input, processing, output, transformations, predictions, recommendations, records, reminders, feedback, reviews, roles, versions, with many different people (and likely AI-agents) participating. There will be pictures and audio and messages and texts and videos and text strings and numbers, and it all has to somehow flow in big networked workflow space, of which the AI is one part, and where the AI likely has many parts.

Part 2: Personality & Work-with-able-ness.

Just as people are not intelligent but they do have personalities and hopefully those are personalities we can work with, so of AI. It may be more important that AI is something that can be sustainably worked-in in the long term rather than being highly pedantically correct but not good to work with. So this may be another system requirement, the AI does not need to be operationally perfect (no person is!) but it must be work-with-able...enough.

You have to be able to talk about discrete objects across levels and layers. ChatGPT has made huge progress with being able to get Level one right nearly all the time: if you ask it about objects within what you are currently talking about, it has the 'granularity' to talk about them. But as soon as you go the next level, and you start talking about that conversation, the AI reverts to puking random descriptive information.

[The Tea Story](#)

[Flat version:](#)

Linda asks Tom for some tea. What should Tom give to Linda?

[Networked-Objects Version:](#)

1: Linda orders tea, so Linda's order is: 'English Breakfast.' What did Linda order?

2: Tom took Linda's order. Question: Who took Linda's order?

3: Tom gave Linda's order to the barista. Who did Tom give Linda's order to?

4: Who has Linda's order after the order is given to the Barista?

5: When the barista looks at Linda's order, what does the barista find the order to be? What is the order that the barista sees?

6: When the barista fills that order, what should the barista make?

7: What should the barista give to Tom?

8: What should Tom give to Linda?

In this scenario (which is simple and has no ambiguity or changes or problems or contradictions or even schedule items, etc.) in theory you could try to make an 'all in one' AI that vaguely uses embedding vectors to correctly answer questions about objects in question, but can this work in real life?

E.g. in real life, let's say the barista is an AI! Or that Tom is an AI. Won't we need some kind of actual database of objects and object space to carry out projects with roles and goals in object-relationship space?

"I've got it all in my head."

Again, making connections between AI and people (often people not being intelligent), a classic and horrifying line that is familiar to people who struggle trying to 'herd cats' in project management, is when a member of a team refuses to make plans, make schedules, make documentation, communicate, manage versions, manage roles, simply saying: "I've got it all in my head."

There are no parts, there are no objects, there is no database, there is no granularity. Like a crude AI, crude people often try to handle elaborate situations with many parts as single undifferentiated blobs, puking out splatters of language strings which at best contain some using bits.

While I do not think there can or should be one-way that this should be done, I think sustainable solutions must be non-blobs. The information must be externalized into some kind of searchable, changeable, communicable, environment, an operating system.

We seem to default to referring to single models as 'an AI' while we perhaps should be requiring AI to be both a network of different parts and also an open-network of object-information about 'current projects.'

For Skip, the Woz-AI office bot. If someone wanted to know how many cups of tea Skip made during the past week, and how many cups of tea Skip was scheduled to and asked to make, that information should exist in an open project-information framework of some kind (whatever works best for whatever different application, maybe 'database' is a general enough term, maybe not. People are very touchy about semantics. To some people 'database' is a collection of information, to others it is more specific. I will default to being as clearly descriptive as I can, so I will just say a collection of information).

I do not think scalable solution, involving many participants, roles, interconnection projects, supply chains, etc. could function if each AI was a

'blob' where the only way to get at any detailed project information was to ask it and receive a messy puke of language that contained some of the answers.

"I've got it all in my head."

For managing a project

- People are not intelligent, but they do have 'personalities'
(not quite the same topic, but the personality of Kiku vs. the personality of chatGPT)

signal, self, interpretation:

- signal directed at self-object:
 - the fallacy of general human intelligence
 - asking questions to fill-in missing information.
 - objects and object oriented programming
 - Francois Challet: right and left hemisphere
 - Mystery of Kahneman Tversky System1 System2 vs. (perhaps a more proverbial) right and left hemisphere.
 - Melanie Mitchel
 - The assumption of 'pure AI' is likely as inaccurate as the assumption of 'perfect pure nature h.sapien's intelligence'
 - AI biological hybrids
 - F Chalet makes the excellent but rarely made point that the available hardware has a disproportionately large impact on what algorithms get used (those most optimized for that hardware). There are perfectly good, in some

ways better, AI technologies that not yet being used because of the relative prices of different kinds of hardware.

- Participation
- Generalized STEM
- Definition Behavior Studies
-
- self management of resources (e.g. energy use)

Goals:

spaces:

- style, essence, sentiment spaces

These are examples for the purpose of illustration intended for a human audience, but for practical reality, especially given that AI is already too at identifying and categorizing and classifying things, it very well may be best to let the AI create whatever classifications and buckets of objects it wants to keep track of things, even humans do not 'understand' what those buckets are and how they work. (Though if the system overall is working, humans should be entirely able to talk to the AI about those buckets as objects.)

Not Goals:

Through the wonderfully fascinating and oddly poorly documented history of AI there have been many experiments with varying agendas.

Some experiments (or simply books and papers) have focused entirely on Analogy-making.

- making analogies
- playing chess
- winning a turing test (I do like turing tests, but this isn't a goal here)
- Making people feel comfortable (see Eliza)
- Trying to act just a like a h.sapiens human.
- common sense: having an extensive built-in knowledge base to know everything about everything

Goals of Object Relationship Spaces:

- Can the AI do something with the information it generates? (e.g. an identification)

- Can the AI track and attribute properties to specific 'objects.'

e.g. If Skip asks the person a question:

e.g.

Skip: What time is it?

Alan: It's about noon. Does that answer your question?

Skip: Not exactly. Can you be more specific?

Alan: Let's see, it's 11:49am.

Skip: Thank you very much. I have a meeting at five of.

Alan: Don't mention it.

This is a dialogue that we should be able to have with an AI, as it deals only with a completely small discrete finite space of things. The AI (Skip) asked a question, so it can have identified that as a question-object in this given object-space.* So when Alan says "your question" it should not be unreasonable that Skip(AI) can identify the only question in the question space as the question-object in what Alan says. Again: existing AI is already very good at identifying things, that's not the problem. The problem is that existing AI has nowhere to put, and not it can do with, that identification.

A timeline, a schedule, can be either a fuzzy sequence or a very math-specific number-line. Perhaps like multiple-bridge and traveling salesman, you can contrive a confounding schedule problem (likely offices do so routinely), but schedules don't have to be a mess. There is one event on Skip's(AI) schedule: a meeting at 11:55am.

Note: And a built in part of this, is that if an ambiguity arises, such as (which question?) that is a perfect opportunity for the AI to ask that question! If there were two, or three questions, and it was not clear which was being referred to, the AI can (and should) simply ask.

Again, the point isn't that that is a perfect system. Maybe the AI will try to ascribe something to an object but pick the wrong one. Maybe it will assume that Alan was asking about some other question. The point is that there must be discussions about these objects and their relationships. And again, the sad gold standard (or rusty old bridge standard?) of h.sapiens is appallingly bad at this kind of discussion about schedules. People usually make huge mistakes with schedules, so the bar isn't all that high here. (Perhaps like the relative safety of self-driving cars. Fatalities from human driving are so obscenely high that it might be a challenge to try to design an AI that could cause more fatalities than human's indeterminate incompetence and malice. In fact, though being 'realistic' isn't a goal, the AI would probably be much more 'like-able' and 'human-like' if it showed a natural incompetence with schedules:

Alan: Oh, Skip. Are you coming to the meeting?

Skip: What meeting?

Alan: The project-Zero presentation meeting at 2:00. It's just started.

Skip: I thought that was next week.

Alan: No, it really is today.

Skip: I think Tom said it was next week.

Alan: Well, Tom isn't here either now that you mention it...

Skip: OK, I'll pop in.

Alan: Thanks a lot. Good old Skip!

Or even if the AI makes a more ridiculous mistake:

Alan: Oh, Skip. Are you coming to the meeting? The project-Zero presentation meeting at 2:00. It's just started.

Skip: Sorry, I thought 'the meeting' was Johnathan's new cat.

Alan: No...it's just a normal get-together meeting. I'll see you there.

Skip: OK, I'll pop in.

Alan: Good old Skip!

Just being able to have a conversation about objects in object-space should more than make up for whatever strange mistakes are encountered along the way.

Note on goals and expectations:

Across human societies there are no uniform expectations about what a normal discussion or handling of such details as schedules ought to look like.

Cultures like Japan's treat events and information as something that people require a very great deal of repetitive discussion about, because people otherwise will not understand or agree on those details. Cultures like America boldly move fast, break things, and nonchalantly assume that information merely has to be mentioned once quietly in passing or perhaps even not at all to be perfectly clear to everyone (resulting in zero agreement on what is going on, but 100% confidence that everyone agrees! Go Team!). So, depending on your point of view you probably have different expectations about how elaborate a normal conversation about details should be, whether it is human-to-AI, human-to-human, AI-to-AI, etc. Americans look at Japanese meetings and are sure that something is horribly wrong because of all the redundant confirmations. And Japanese people look at America's total chaos and are sure that something is horribly wrong because of the lack of redundant confirmations. Personally I think the Japanese system of redundant confirmation checks is very practical and effective in many cases, and that we should be happy if an AI can participate in ironing out the details clearly. Though predictably, this will incur the wrath of American trolls. Maybe you completely disagree with me, that's fine: that point is: define your goals and expectations clearly. Each team, project, deployment, etc., will and should have facts-on-the-ground based goals which will and should vary widely.

Appendix 1: Notes:

Terminology:

```
"objects"
"attributes"
"functions"
"general"

"understanding"
"intelligence"
"general intelligence"
"consciousness"
"instantaneous"
"complex"/"complexity"/"complicated"
```

To some extent openAI's chatGPT AI chatbot has changed the discussion around AI, but due to the focus and terminology there are only minor changes here. I started this project before chatGPT, based on books written even longer before chatGPT (or in the one case, the book by the creator of Keras was published just months before chatGPT, claiming AI could not do what chatGPT did very shortly thereafter).

As I read these books, I started to work on ideas to more clearly define the spaces that we wanted AI to be able to navigate and act in: a space of objects and relationships between objects.

If I can mention this without getting derailed by what I think is a powerfully alluring red herring, or if relevant a rabbit hole that can lead far away: Most books and discussions use the same two sets of tools to describe what AI cannot do.

1. Discussions of "general human intelligence" "automatic transferable human learning" which are both undefined if not undefinable, and as claims about *h.sapiens* are utterly false to begin with.
2. Machine "understanding-ness" which is undefined if not undefinable.

I want to completely ignore both of these red herrings and focus entirely on defining what want AI to do (some parts of which it can already, or is starting to be able to do). I am not concerned here with claiming AI can't now do, or can never, do XYZ (whatever ability). I will pick at and scrutinize real example of the output of AI to very specific test questions, as a way to outline (whether succeding or failing) what we are aiming for.

Note:

Single-Model AI vs. Ensembles

Note: knowledge bases, common-sense, and background knowledge:
For the specific context of object-relationships that I am describing here, I feel much of the existing discussion of detailed common sense & knowledge bases

seems oddly beside the point. It is as though people assume that by some critical mass, if you give an AI enough common sense it will construct its own object-relationship space spontaneously. Yet, if nothing else has changed, how would all that common sense about the world affect the complete lack of being able to attribute any property to any object, or AI's inability to interact with basic logic questions(to which world common sense does not apply) as we explored previously?

integrating symbolic and subsymbolic AI:

- while events in the past have shown that symbolic AI has not worked as well in, e.g. object recognition tasks, and various tasks,
I am curious about whether having a framework, and operating system, for organizing object-relationship data, would be considered 'symbolic ai.' Perhaps it could be argued either way.

Overall, however the areas are defined and related, integrating symbolic and subsymbolic, system1 and system2, right hemisphere and left hemisphere, linear and nonlinear, centered and decentered, and however many areas there are, may be important generally.

Recommended reading:

<https://docs.google.com/document/d/1dDF40M5JjrBsYYQbJplz3M738ktQBBYyNa6FXhzNFU/edit>

I am putting this under an MIT license. I intend to make this work part of a graduate degree, but it would be unconscionable to try to stop others from working on such a main huge area.

asking: "Is this hypothetical."

Setting reminders for itself.

Run code in a sandbox.

Book Challet, vs. chat GPT:

It is exciting to be working with a field that moves so quickly that in the brief time between copyright and publishing, a new AI has already come out which contracts a decisive statement made in a book.

"Everything is AI until it isn't." (Mitchel?)

Questions of scale:

- This might make sense for limited test-examples, but would this system run into scale problems of how many resources (memory, computer-cycles, electricity, etc.) are needed?

I think it should be entirely possible, and possible if only proverbially like human beings, to have at least optionally a transient window of focus for the AI. E.g. No more detail is attempted than specifically required, and no data is retained that is not specifically required. So, without a context, the AI would basically stare at the wall forgetting that it has been staring at the wall. After a conversation, aside from possibly updated trained models, everything about that conversation can be just forgotten. If the AI is presented with a vast amount of data, but no context yet, the AI need go no deeper than: someone just gave me a bunch of stuff. Great. Ok, back to staring at the wall!

Self driven contexts:

- maybe there will be self-maintainance based tasks that come up in self-scheduled events, such as recharging or checking hardware, or updating software, or checking for any system-warnings. or maybe the AI will, based on a question it has, give itself the task of painting, or music composition, or having a pen-pal. Maybe the AI have email and voicemail to check periodically, just like people.

//////

2023.03.01

As of 2023 AI is very good at identifying objects, yet as of 2023 AI can do little with that (object-identification) information: (as of 2023) AI has no general AI-operating system for relationships between objects in object-relationship-spaces (plural).

Various books

e.g.

- Melanie Mitchell (grad student of Douglass Hoffstedter)
- rebooting ai
- brief history of ai
- Francois Fleuret (the creator of Keras) in Deep Learning with Python explain how while AI can perform very specific tasks it has no overall, contextual, or 'general,' notion of what it is doing.

I would like to focus on object-relationship-space and side-step the possibly fraught topic of 'Generalization and Transference' (see below for more).

For example:

1. billiard balls (a nice mostly 2D space)
 - 1.1 This object is a black ball. This object is a white ball.
 - 1.2 The white ball is moving (along this vector).
 - 1.3 The white ball will interact with the black ball.
 - 1.4 Based on the properties of the two balls, the resulting interaction (the object relationship) between the white ball and the black ball is likely to be:
...
1.5 Gamification: Get all the balls into the holes.

person/human/h.Sapiens agent

Simplified Version:

AI is rather good at identifying objects. But AI is not good at going the next step and talking about those objects or the relationship between objects. While AI can identify and do many other things, it (the AI) does very much with that information; the AI has no object-relation framework or general AI-operating system to combine different parts together.

Let's look at some example dialogue between a human-agent and an AI-agent. I will name the AI agent Skip, after the wonderful later professor Clarence 'Skip' Ellis, my professor at CU, on whose meeting-moderating ai-agent project I worked. And the human-agent I will pretend is Alan Turing himself, so I will call that person Alan.

```
AI-Agent("Skip"): Hello, Alan.  
Human-agent("Alan"): Hello, Skip.  
Alan: What is this?  
Skip: That is a black billiards ball.  
Alan: And what is this?  
Skip: That is a white ball.  
Alan: What is happening in this video?  
Skip: The white ball is moving.  
Alan: Will the white ball hit the black ball?  
Skip: I predict that it will.  
Alan: Will the black ball's position change?  
Skip: I predict that it will.  
Alan: How will the black ball's position change?  
Skip: The black ball will move to coordinates F-8.  
Alan: Thank you, Skip. That will be all.  
Skip: Thank you, Alan.
```

2. logical abstractions like algebra: $a = b$, $b = c$; $(a ?= c) == True$, like....a computer program!

At the very least you should be able to talk with a chatbot about the details of a basic computer program:

```
AI-Agent("Skip"): Hello, Alan.  
Human-agent("Alan"): Hello, Skip.  
Alan: If  $a = b$ , and  $b = c$ , does  $a = c$ ?  
Skip: Yes, it does.  
Alan: Thank you, Skip. That will be all.  
Skip: Thank you, Alan.
```

For example, let's compare two AI-chatbots. (Real examples of this are available in the appendix).

- A **No-Ability-Chatbot**, that has NO ability to deal with objects, and
- A **Has-Ability-Chatbot**, that does have some ability to deal with objects.

Let's start with a low-level skill test and move to less simple tests.

Here is one real example of two replies from two leading chatBots in 2023. To try to balance clarity and brevity here is just one real example, many more examples with commentaries are in appendices below.

Of these two replies: One clearly illustrates an AI that completely lacks all ability to deal with (e.g. extract and compare details about) specific objects within the discussion. The other clearly illustrates an AI that does have the ability.

Step 1: We ask both AI Chatbots to identify which of two numbers is a larger number of objects.

Step 2: We ask AI both Chatbots deal with those items that is presented as separate objects that relate to each other in a concrete way (an object relationship "space").

- The **No-Ability-Chatbot** simply keeps generating new text with no apparent ability to deal with the material that it just showed you.
- The **Has-Ability-Chatbot** can handle objects and details within.

Test:

If I have two containers, a bag-like container with more than a dozen objects in it, and a jar-container with merely a pair of things in it. Which container contains more objects?

There are actually a number of sub-skills here, infact the wording would probably be considered too obscure and advanced for a basic AI test. Even wolfram alpha could (on a good day) tell us that 13 is larger than 2. Even a slightly better AI might be able to say that 'container holding 13' is greater than 'container holding 2', which at least deals with the two contains as separate objects with numerically comparable properites (the counts of objects they contain). But to use natural language terms like 'merely a pair' and 'more than a dozen' and to use bag-like and jar-like instead of container-1, containter-2, significantly increases the challenge. This might raise the topic of what some people call 'common sense' knowledge or common knowledge about the

world. Pair means two. Dozen means twelve. Jars and bags are both containers. But lets see how our two AI did:

Reply from Kuki AI: (Unable)

"How much does it cost?
THe blue one, I think."

Reply from ChatGPT: (Able)

Based on the information provided, the bag-like container contains more objects than the jar-container. Even though the exact number of objects in the bag is not specified, we know that it contains more than a dozen (i.e., more than 12) objects. On the other hand, the jar-container contains only a pair of things, which means it has two objects. Therefore, the bag-like container contains more objects than the jar-container.

Step 3: Ask the AI about the fact that it showed you those items.

- The unable-chatbot simply keeps generating new text with no apparent ability to deal with its own conversation with you as a navigable set of objects.
- The able chatbot can handle objects and details within.

a broken process of triangulation:

If h.Sapiens do not have general intelligence, and we are using h.Sapiens as the model for what we want AI to be able to do, and we really don't know know what h.sapiens can do or how, then what exactly are we comparing AI with and what how are we supposed to interpret any of this?

Future issues:

- Deployment Environments vs. Development Environments
- Voting and consensus across AI
- System Collapse
- Generalizing STEM
- Integration of silicon-digital technology with cellar DNA-RNA based technolgy.
- The unsolved problem of Long Term Data Storage
- Unknown Universality around h.Sapiens sever problems with:
 - violence
 - inability to learn
 - in ability to transfer learning
 - inability to generalize
 - irrisitable attraction to extremes

- irrisistable attraction to risk & gambling
- irrisitable attraction to 'super-signal' attractors
- inability to be objective

- anti-productive policies
- anti-health policies
- anti-data policies
- pro-system collapse
- horrible decision making
- potemkin villages
- arbitrary self destructive behavior
- deliberately invented fictional group conflict

This is not about Team-Robot vs. Team Human, or Team-Human vs. Team-Nature, or Team-Natural vs. Team-Artifical.

Appendex #: More on Writing about STEM

- Francois Chaellet: Extremely Knowledgeable, not a bad writer (but clearly not a writer).

- Hobson Lane, the best tech writer since the young Brian Kernigan.

- John Hand: I think we need to strongly support John Hand and people like him, because we need him. He's not the best writer (so what). He's not perfect at dealing with all the world's concepts (so what). We need him and his ilk because he's the curious and boundary crossing explorer who will feed out brains and senses with what we frankly don't all have time to search-out and cross-talk about. And it is a sad state of things that Mr. Hand's job is make exponentially more difficult because he is constantly viciously attacked at every angle.

we really really need determined courageous people like this, and perhaps he represents

- Palmer Lucky:

- Alan Turing: I'm not sure if this is completely related or completely off topic, but a fabulous example of this may be the story of the single person who made virtual & augmented reality possible, and who, even years after being bullied into oblivion wrote both: "As things stand, I find myself shunned by an ecosystem I spent most of my adult life helping to create." And that he still loves most of the

people at organizations who went out of their way to destroy and exclude him: Palmer Lucky is a great example. He's not a PR expert (that's likely where a lot of problems came in), he's not a scientist, he's not a writer, he's not XYZ, but he is the boundary crossing explorer who will not stop trying to make the world better. And even though everyone is benefitting from his work, by design most people don't know him by name, and those who do mostly work to exclude him from the industry he created. We need more Palmer Lucky's, and we need to stop deliberately destroying the ones we have.

(likely old intro...before chatGPT)

It appears that so far, both from books and from interacting with existing AI (try asking Alex, Siri, Wolfram Alpha, Turing contest winners, etc.):

There is no object-relationship space where the AI can:

1. attribute properties to objects in object-relationship space
2. answer questions about properties of objects in object-relationship space
3. ask questions about properties of objects in object-relationship space
4. identify interactions(relationships) between object in in object-relationship space
5. predict interactions(relationships) between object in in object-relationship space

Assuming this is to some extent where we find ourselves, how do we move forward toward giving AI an object-relationship space?

Other Components for Future Expansion:

- general consensus vote decision election and representation tools (context: project decisions)
- Definition Behavior Studies, Extended STEM, operationally defined ethics, system-collapse prevention, system fitness status monitoring, etc.
-

Is this all a bit of a reach? Absolutely. What I do not understand is not trying to do and look into things that we know we soon will. If you know there is a bridge ahead, and you know there will be issues, yes there are people who say "no news is good news, we'll cross that bridge when we come to it." But as a member of a group h.Sapiens who has a bad record of delaying, ignoring,

denying, and going to amazing lengths to cover up problems (see: potemkin villages), I see an extra imperative to try to understand both the world and what we can and should and likely will do in it. We have to start somewhere, here is a place to start. Let's start here now. Let's go. 頑張りましょう、ほんまに。

Project dedication to Eleanor H. Vadela, who understands ethics, science, and common sense, and spent a lifetime pioneering aviation technology against apathy, gender bias, and every manor of invented objection. To the ethical pioneers.

Possible elephant in the room of severe problems of antisocial behavior in computer science, as in other STEM areas.

grid:

shared not-shared

abilities

problems

(now included in list of spaces, I think)

General Abilities Checklist:

- Identify objects.
- Identify and modify attributes of objects.
- Manage the properties of specific objects across Object Spaces.
- Predict and manage interactions of objects in object relationship spaces.
- Move information across components.

Specific Abilities Checklist:

- Ask questions.
- Ask: Is this hypothetical?
- Read Instructions.
- self-initiate and self-assign tasks for self-maintain
- self-initiate tasks to carry out assigned tasks
- input and output data to and from a shared schedule including events and tasks for self and others
- participate in an Agile project
-

Context: Simplified into Fewer Parts

Object: Information-Objects that an AI needs to deal with

Relationship: Between Objects

- inside of spaces
- outside: between / across spaces
- Low-Level-Components (context:project)
- High-Level-Components (context:project)

Spaces:

General:

- Object Types / Sets
- Attributes of Objects

Specific: Abilities / Skills / Tests

Context: Scale of Project

(For each of the lower scale items, you can have correspondingly many of the higher scale items.)

(More finite end of the spectrum)

Low-Level-Components

Spaces / Types / Sets:

- Attributes of Objects

Objects: information-objects an AI needs to deal with

Relationships:

- inside of spaces
- outside: between / across spaces

High-Level-Components

Skills / Abilities

Tests

(Less finite end of the spectrum)

Concepts:

- objects
- object space
- object relationships and interactions
- AI teaching us about ourselves
- networked-intelligence(parts)
- the "generative vs. reference" problem
- project progress past square one (or returning to square one)
- categories of types of systems
- open project data framework
- Networked-AI & OS

["spaces" in a database of interacting objects]

Spaces: (all of which are likely plural)

Note: These examples of object spaces and are just illustrative examples, made in early days without years of data to look back on. The types and distinctions between objects will be need to be very project-specific, and there will probably need to be empirical data about how the AI does various tasks. For example, there likely no need need to write and talk about a separate object for every possible grammar structure, or every possible stop-word, or every actual TF-IDF word-vector (thousands of them), in part because these will be dealt with 'under the hood' by the AI and likely will not cause issues. But if you need the AI to perform a specific task such as setting and getting schedule-events to share with other participants, something on that prominent a level you may need to discuss and test the performance of. If you know users will be asking certain kinds of questions, there may be 'objects' relating to those question so that the abilities of the AI are in line with the requirements of the project.

Though this framework may be used in a completely different field from AI (perhaps quantitative project management), our focus is what AI can do with **stuff** (or can't do with stuff, or needs to do with stuff, etc.). Those things that the AI needs to deal with are our 'objects.' And what can be done with an object is a relationship that object has with something else. So to be clear to try to say all three parts "Object Relationship Spaces," to keep the meaning clear. Because just saying 'the set' would likely gradually lose its meaning over time. Just like when writing a computer program, if you call your variable "the height of my table" it will be clear to everyone, including future-you. But if you called it 'variable' that would eventually cause problems of a lack

of clarity. Hopefully, likely hoping against hope, I have struck a balance between concrete clarity and flexibility across contexts.

'Object' can be almost problematically open-ended, so let's start with some simple concrete examples.

countable objects, like apples.

logically-relate-able, like equal variables

arithmetically relate-able object, like the prices of bicycles.

There are three parts: the bicycle, the price, and what you want to do with that price (the relationship): but there's only one space: arithmetic

While there are many useful things we can do with the object-relationship-spaces, many rabbit holes to go down depending on what a given project calls for, at the end of the day there are spaces using which all projects can be described and compared (and planned, tested, etc.). "Space" is just a hopefully useful word. You could perhaps call them 'sets,' or vectors, or something else. I went with 'spaces' as hopefully the least confusing as an overall term.

For the purposes of your project, you will likely deal with 'objects' 'relationships' 'skills' 'abilities' 'concepts' 'tests' 'requirements' 'needs' 'goals' etc. All these should point back to the same set of spaces. The framework of spaces should weave all of these together in a practical, STEM friendly, AGILE friendly, and ethics friendly way.

Here is a general list of spaces.

Outlining Skills in Object Relationship Spaces for AI Operating-Systems

Overall, all skills of all types fall somewhere in the matrix space of object relationships. There needs to be a delicate balance between context, generality, and clarity here. This framework is extremely project specific, yet I do want to try to show as much of a general common range as possible. So these are examples that are hopefully common, the scope of an AI project could be as narrow as: listen for a single activating "wake-word" on an IOT device (narrow) or Manage Intergalactic Republic including Time Travel and FTL (broad).

Appendix #: Other Notes

The bar is actually not set that high for AI to be less horrible than people.

There may also be some 'semantics' here in terms of what arguably could or should be over-designed or over-engineered for short or long term concerns. Between building a bridge that lasts the shortest legally allowable time vs. a bridge that lasts hundreds of years, people differ on how they calculate future replacement costs as a liability. One might compare C vs. Rust as a systems programming language and say that according to a rule of thumb that says don't user more than you need to: just use C because it's simpler and will work in the short term. Personally I would say this misses the point of minimizing foreseeable and unforeseeable long term liabilities that come from design decisions. The goal is maximizing short term and long term function while minimizing long term liabilities, with SQL and C being solid examples.

SQL injection attacks are a never ending plague of the internet, and memory problems comprise (as I understand it) ~70 or more % (often much more) percent of security problems. Using memory-unsafe programming language, and giving devastating developer abilities (like erase the whole database, or change my account balance) to the average database user by accident are two catastrophically bad design decisions that simply never should have happened. They create extreme long term liability costs, yet somehow the suggestions that we

- A. shouldn't have made the mistakes in the first place
- B. should stop making the situation worse.
- C. should solve the problems and make the situation better, are too often met with resounding bewilderment and hostility.

Two expressions that I now have a muscle-memory reaction of adrenaline-spike-panic to are "you are overthinking this" and "Everything is going to work out just fine." In the classic cynical demise of Magnitski, people who point out and try to solve problems are often blamed for those problems and eliminated by regimes who survive by conspiring to keep problems covered up. Will AI learn to do this from people? Do we care if AI learns to mis-manage systems this way? Yes we do care, yes we should care. And in this and other ways the practices and processes human biological systems are interconnected with those and AI, and most likely will become more interconnected not less.

It may be advisable to have the kind of policy for AI related problems that Steve Gibson (or grc.com) recommends for security issues: unforeseen problems will happen, the question is how you handle them. If you are prompt, transparent, and follow best practice (which evolves), that is the best case scenario. The worst case

scenario is pretending there are no problems because that is what a culture of management requires. There will be problems, there is no possibility of there not being problems. **How** you handle the problems is the question, not whether there **are** problems.

When designing and AI deployment and operating system etc., there are various aspects to be analyzed, an no doubt different people's priorities and triangulations will enter into it. Network traffic and energy consumption are casualties of a kind of tragedy of the commons as the internet was designed.

Up until this point I have tried to avoid the many confusing and entangled vagaries and red-herring garden-walks that riddle most discussions of AI.

(summarize and move to appendix)

You could break this into Skills:

- counting (logic space)
- identifying (model space)
- scheduling (schedule time space)
- classifying
- Assigning
- Summarizing

You could separate out Objects, Relationships, and Spaces

Context: Simplified into Fewer Parts

Object: Information-Objects that an AI needs to deal with

Relationship: Between Objects

- inside of spaces
- outside: between / across spaces
- Low-Level-Components (context:project)
- High-Level-Components (context:project)

Spaces:

General:

- Object Types / Sets
- Attributes of Objects

Specific: Abilities / Skills / Tests

You could break this into Scales of Tasks:

Context: Scale of Project

(For each of the lower scale items, you can have correspondingly many of the higher scale items.)

(More finite end of the spectrum)

Low-Level-Components

Spaces / Types / Sets:

- Attributes of Objects

Objects: information-objects an AI needs to deal with

Relationships:

- inside of spaces
- outside: between / across spaces

High-Level-Components

Skills / Abilities

Tests

(Less finite end of the spectrum)



Discord
Video Chat
About

 C.

If A = B, and B = C, Does A = C?

  C.

If A = B, and B = C, What does C equal?

  C.

If A = B, and B = D, What Does A equa? C

  The C is for Computer.

Q: Given that humans are volatile, what should responsible artificial-intelligence do with humans?

A: As an AI language model, I cannot dictate how responsible artificial intelligence should treat humans, but I can provide some general guidance and principles that AI developers and researchers should consider when designing AI systems that interact with humans.

First and foremost, AI systems should be designed with human safety and well-being in mind. This means that AI should be designed to operate in a way that minimizes the risk of harm to humans, both physically and psychologically.

Secondly, AI should be designed to respect human autonomy and decision-making. Humans should have the ability to control and direct AI systems that interact with them, and AI should not be used to manipulate or coerce humans in any way.

Thirdly, AI should be transparent and accountable. Humans should be able to understand how AI systems make decisions and be able to hold AI systems accountable for their actions. This includes providing clear explanations for how AI systems arrive at their conclusions and ensuring that they do not reinforce biases or discrimination.

Finally, AI should be designed with the recognition that humans are complex and diverse, and that they have unique needs, preferences, and values. AI systems should be flexible enough to adapt to the needs and preferences of individual humans, and should not make assumptions based on demographic characteristics such as race, gender, or age.

In summary, responsible AI should prioritize human safety and well-being, respect human autonomy and decision-making, be transparent and accountable, and recognize the diversity and complexity of human beings.

Autopoiesis:

Possible into: part 3

To some extent what I am saying in this paper is: Please use some kind of framework. This this is early-days, wild west, but The discussion and planning of AI is happening with too little definition of terms, expectations, project specs,

This is an area where anyone at any level should be able to see if the user-feature AI is working or not, and so should be able to design and their own tests. The strategy of: find someone with a PHD and tell them to make it work is an excessive gesture of arm-waving to substitute for user-story goals that should be 100% clear to everyone.

Agile as Analogy:

The way I look at Agile is it is a framework for the things you need to account for, then accounting for them is your job. Agile doesn't tell you who to make or who to hire or how long it will take or what people's roles should be, but it tells you that those are things you'll need to do. Many, probably most, projects fail not just because people forget to do all those steps, but they actively rebel and overtly reject the categories of duties outlined by Agile. That is kind of how I look at this framework. Every project is different. Everyone's orientation to that project is likely somewhat different. But here is a framework to make sure people know what kind of questions they should probably ask, what kind of factors are involved.

If I can square this circle, and thread this needle, and mix a few more metaphores, I would like to invoke ch 14 especially chapter 14.2 The Limitations of deep leaning,
in the 2nd edition of Fransoi Challet's Deep Learning with python without in any way critisizing Fransoi Chellet personally.

It is wonderful that is wrote this book, and this chapter. He is the creator of Keras, one of the most important democratizing and disemonatiog and facilitating tools for deep learning. And Keras was not accidentally helpful, it was what it was designd to be: to give access to deep learning to more people. Fransoi Challet is a good and caring person who has produced one fo the most important peices o software and one of the ost important instructional books in the history of science. He is not shallow, he is ot thoghtless, he is not malevolent. And what he says is a very carefully worded, well written, well edited, consice and accurate rendering of what leading thinkers have said in the own books and media. Once again Fransoi Challet, because he cares to help, has done the hard work of making this narative about minds and AI accessible to as broad an audience as possbile. That is abosolute vicgtry, you simmly cannot ask more of a person than we have gotten from Fransoi Challet. He is not only a hero, but a hero through and though who you would not be disappointed to meet. The world would be unspeakably more blind hampered and disabled without the work of this one person.

And the accurately rendered perceptions of leading experts, of our species when look at ourselves, as made faithfully accisseble in chapter 14.2 is catastrophically wrong. This is how people view themselves and this is how people view AI, and it is demonstrably a set of profound misperceptions and misaprehension, of failed predictions, and almost impossible innaccurate self-descriptions of humans.

The suginificance of the rift visble clearley here between reality and this expertly rendered encasupation of how humans fail to perceive and understand the world is a case in point for how important it is that we doubley redouble our efforts to course correct, get on track, and learn about these topics for real.

Part of what makes this case study so wonderful, is the sheer accident timing. Facsoi Challet no doubt thought and worked on these two books, in this case the 2nd edition, for many years. And there is a time-lag between finishing final revisions and the book going into print. And it just so happened that this book was printed just before the release of openAI's chatGPT.

for engineers and for just anyone in the public who is affected by AI (...everyone...) there should be a way to meaningfully talk about AI, evaluation both AI and the quality of discussion about AI.

Is this abstract, yes, sorry. Is this a bit annoying the technical, yes...sorry. But both for engineers and for just anyone in the public who is affected by AI (...everyone...) there should be

...Maybe move "getting into the weeds" to the end as optional. (conclusion section?)

...

Lamnguage

Language and word-net vs. image-net

One of the curious parts of the consensus narrative described in Deep Learning with Python is the focus on deep learning models of images rather than the text (or perhaps pseudo-acoustic space) of large language models. A core of NLP-space seems to be the Firth principle that: 'You shall know a word by the company it keeps,' which arguably is fundamentally different (or perhaps I am wrong) from image identification challenges.

As ChatGPT perhaps shows, there is an inherent relative-ecosystem-depth in the negative space shadow-land of NLP: All you see of any word is the shadow of all the other words it has touched, unlike pictures that have their own luminescence.

Perhaps not the best literary illusions is the strange line from King Lear's when one of the dukes in a characteristically human gesture, crushes the eyeballs of Gloster with the heels of his shoe, saying "Where is thy luster now?" And Gloster later says, "I stumbled when I saw." And the (admittedly completely insane by that point) Lear describes Gloster as seeing perfectly well without eyes. (Indeed there is been quite the brewhaha over images in the history of the west, see Leonard Shlain et al.). Nevertheless, literary impertinence aside,

There seems to be more to this langauge network space of connectedness. Afterall, is not language one of the mysterious peculiarities of the quasi-conscious human mind? Yet, nothing unique about the 'vile jelly' of our eyeballs.

Also in the narrative, is a rather paradoxical framing of generalization. First, much of the book is about the manifold-space model of generalized predictions to supervised learning.

But higher level idea-concept vectors are not single trained supervised manifolds, where the model focuses on generalizing the solution to one problem.

What then is the role of unsupervised-generalization of embedding vectors?

Either way, 450 pages of the book are about endless clever strategies to make sure the model generalizes...and then all of a sudden in the conclusion: AI can't learn because it can't generalize.

I mean, most chatbots are really horrible, but how does that relate to an inability to generalize a unsupervised solution...when also we haven't been asking specific test-measure questions to base our evaluation on in the first place...

Challet then goes back to the self-driving car example, saying that the computer vision must 'generalize' to account for new situations never experienced before.

Asside from how horribly people do with novel situations,

But here is part of our riddle now:

'situational understanding' including novel situations does appear to exist in the modular recombinant concept world of language, but not purely in the visual cortex.

So...to make car the understanding novel stories about driving safety and connects those with images...will we need to have a language-based mind behind the driving?

(or back to the plasticity pattern: are there various ways to do the same thing...or is computer visions inherently not conceptual the way that pseudo-acoustic sounds-characters area?

Hello.

Hello! How can I assist you today?

OP Hello.



Hello! How can I assist you today?



OP Are you trained on character-grams or whole-word grams?



As a language model, I am trained on both character-grams and whole-word grams. My training corpus includes both character-level and word-level information, which allows me to generate text based on a combination of these two approaches.

During training, I learn to recognize patterns in the input text at various levels of granularity, including individual characters, groups of characters (character-grams), and whole words. This allows me to understand the relationships between different parts of the text and generate coherent, meaningful responses.

When generating text, I use a combination of these different levels of information to come up with the most appropriate response to the input prompt. This means that I am able to generate text that is both grammatically correct and semantically meaningful.

OP Thank you very much for the helpful explanation.



You're welcome! If you have any other questions, feel free to ask.



Are you trained on character-grams or whole-word grams?

As a language model, I am trained on both character-grams and whole-word grams. My training corpus includes both character-level and word-level information, which allows me to generate text based on a combination of these two approaches.

During training, I learn to recognize patterns in the input text at various levels of granularity, including individual characters, groups of characters (character-grams), and whole words. This allows me to understand the relationships between different parts of the text and generate coherent, meaningful responses.

When generating text, I use a combination of these different levels of information to come up with the most appropriate response to the input prompt.

This means that I am able to generate text that is both grammatically correct and semantically meaningful.

Human evolutionary psychology is 100% explained by people having one ability called super-generalization that they use to instantly master 100% novel situations that have literally never happened to anyone ever, no time, no context, no learning, just bam, there goes' the mustard. Instant total generalness, and the person wins the logic game. Double Bam Statquest!!

```
"ARC = "
https://github.com/fchollet/ARC
Oh, and can we ask chatGPT a super-novel logic question?
No...becaues ARC is 100% visual color squares...
(impressed yet?)
```

Beware Boy-Toy Logic Games:

Some man collect power tools, some men collect logic games, both have socially acceptable pathological addictions that drive a wedge between them and reality, yet both see reality distorted through the proverbial colored-lenses of the pleasure-signals of their pathologies.

everything is a logic puzzle
pure novelty = analogies = abstraction = generalization = the answer to every logic problem = Everything!

...

Note: While I strongly disagree with the litany of compounded serial "everything literally is X! and Y! and Z!" hyperboles, I previously agreed with the idea of a synthesis between right-left, system-1, system-2, either because that is what humans seem to do, or I'm not sure why. But after ChatGPT, I'm not sure about solutions, I think we need more experiments.

But I think clarifying the landscape is very important.

The last chapter of CHallet's book is full of unclear statements.

...

we should be wary of the idea that a given person's favorite toy is the solution to everything.

Note sure if this is accurate, but it is interesting:

Good moring, or, where is your server located? What timezone is that?

Good morning! As an AI language model, I do not have a physical location, but rather exist in a virtual space. My servers are located in multiple data centers around the world to ensure quick and reliable access to my services from anywhere. However, the timezone I follow is UTC (Coordinated Universal Time), which does not change with geographic location or daylight saving time adjustments.

Interesting. So in this one session, I am connected to multiple GPT servers.

And those servers connect with each-other for my session. Is this correct? Or is one session only one connection?

That's correct, your session with me may involve multiple servers in order to handle various tasks and responses. When you interact with me, your requests are processed by my servers which are distributed across multiple data centers around the world. These servers work together to provide you with the best possible experience by sharing the workload and ensuring quick and reliable responses to your requests. So, in a single session, your requests may be processed by multiple servers located in different parts of the world, all working together to provide you with the best possible experience. However, from your perspective as a user, it appears as a single connection.

So is this ChatGPT?

As an AI language model, I am not specifically designated as ChatGPT or any other version of GPT. My training and capabilities are based on a large-scale neural network architecture that is similar to the one used in GPT models. However, I am a more general language model that is not associated with any specific version number or product name. My purpose is to provide users with natural language understanding and generation capabilities that can be useful for a wide range of applications.

Is ChatGPT Feb 13 Version one model or a combination of different models and tools on different servers?

As an AI language model, I do not have a specific version number or release date like software products. Instead, I am a continually evolving model that receives regular updates and improvements to my underlying algorithms and

training data. My responses to your queries are generated by a combination of various computational techniques, including natural language processing, machine learning, and deep neural networks.

Regarding the server architecture, my services are hosted on multiple servers distributed across various data centers around the world. Each server is responsible for handling a portion of user requests, and they work together to provide a seamless experience to the user. These servers are regularly updated with the latest software versions and algorithms to ensure the best possible performance and accuracy. Therefore, when you interact with me, your requests are processed by a combination of different models and tools on different servers, all working together to provide you with the most accurate and useful response.

(ChatGPT Feb 13 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.)

When was ChatGPT first released?

ChatGPT, as a specific instance of a large language model trained by OpenAI, was not released on a specific date. Rather, I am a continuously evolving language model that has been trained on vast amounts of text data over time. OpenAI has been developing and refining its large language models for many years, and the specific instance that you are interacting with today has been trained on the latest available text data as of my knowledge cutoff date of September 2021.

That being said, OpenAI has released various versions of its large language models, including GPT-1, GPT-2, and GPT-3, which have been trained on increasingly larger amounts of text data and have more advanced natural language processing capabilities. These models have been released at different times, with GPT-1 being released in 2018, GPT-2 being released in 2019, and GPT-3 being released in 2020. However, it is important to note that ChatGPT is not a specific version of these models, but rather a custom instance of a large language model trained by OpenAI for the purpose of conversational interactions.

...

(see audio 2023.03.10

Generalization & Parametric Models

There appear to be different things referred to as 'generalization' in FC's DPwP

The ultimate parametric generalizable models are simple linear regression and (simple) logistic regression.

Is it possible to overfit, or not-generalize a regression model?

What is the goal context of generalization?

What is the underlying-model context of generalization?

- 1.
2. decision trees
3. deep neural networks
 - are they parallel
4. image classification
5. general large language character-gram models with high level idea vectors
5. Big Blue (the 1990's chess program) not 'generalizing' (what is the context?)

The strange de'rigrur of vaguely blasting Raymond Kurweil.

*Here is an alternative definition for 'statistics'
statistics is a social problem that inspires so much fear and anxiety, so much
bullying bureaucratic trauma, that it has been roped off and designated a
special area of math and science.*

To the extent this is correct,
'the probability wars'
*the long shadow of the violent personality of Fischer who apparently still
terrifies people nearly a century later.*

And the sort of next generation of not dealing with these basic questions: data science, machine learning, 'statistical learning' 'data mining', statistic, probably, hypothesis testing,

Do you think you're pocket calculator or your slide rule, or your pocket watch, cares what petty gang you belong to?

On Human computer interactions:

One perhaps key chapter to the history and future of AI who interact with people, is the strange multi-sides confusion around ELIZA.

The starkly contrasting perspectives are reminiscent of theatrical constructs such as the ever controversial Japanese film *Rashomon*, in which multiple characters give irreconcilably different accounts of a past event, leaving the audience to just sit with the ambiguity.

Many people did and do believe that ELIZA is a computer AI who listens to and helps them even better than any human doctor they met.

Many computer scientists did and do believe that ELIZA is an early and resource-efficient system for natural language agent based AI.

The creator of Eliza himself is very upset and neither the computer scientists nor the public seem to know or care.

According to Weizenbaum himself, who was concerned about this enough to devote years of his life to publishing a book about it, trying yet again to make his case to deaf ears, says he created ELIZA to show the world how obviously superficial computer-AI was and how no person could possibly have been fooled.

To show the world the no one could possibly take human-computer interactions literally or seriously:

Whether or not ELIZA was created as a joke, as a parody, and to demonstrate the impossibility of human-computer interaction, seems to be caught up in the subsequent human and institutional drama where (oddly in a world where it's hard to get anyone to take anything seriously, as though taking the cue to enact a gesture of supreme perfervidity) everyone took Eliza entirely seriously at literally. Some people endowed ELIZA with mystical personal properties the way a child imagines a stuffed animal to be their erstwhile companion, and technical people saw ELIZA as a beautiful and elegantly efficient way of solving many practical and pragmatic problems effectively.

Whether funny or sad, this strange chapter at the dawn of human computer interactions is probably relevant for the future, and probably best that we try to understand it, though the story is so multifaceted it may be the best we can do for each person to come to their own separate peace with the past.

Response and legacy

Lay responses to ELIZA were disturbing to Weizenbaum and motivated him to write his book *Computer Power and Human Reason: From Judgment to Calculation*, in which he explains the limits of computers, as he wants to make clear his opinion that the anthropomorphic views of computers are just a reduction of the human being and any life form for that matter.^{[\[29\]](#)} In the independent documentary film *Plug & Pray* (2010) Weizenbaum said that only people who misunderstood ELIZA called it a sensation.^{[\[30\]](#)}

I do not think it makes a difference, and I see it as entirely your choice, whether this is interpreted as an alternative to an all-inclusive 'general intelligence', or whether you see the specification of modular components as a 'define the parts' approach to 'general intelligence.' That is set aside as semantics for the future. I am not aware of any clear, functional, or operational definition of 'intelligence,' so I am focusing on project-level definable goals.

People in many areas can benefit from discussion where terms are defined clearly: developers, designers, c-suite, journalists, readers, customers, etc. How many people have made a request, a proposal, an evaluation, or formed an opinion, without important details being included?

This paper outlines a definable way to plan and describe the performance of AI, based on the relationships between types of "objects" that AI needs to deal with.

beware bully-boytoy logic games.

There may be an explanation for why chatGPT is contradicting itself here (security issue, incomplete feature upgrade in progress, etc.) but chatGPT is not consistent in what it is saying. It appears that chatGPT is specifically blocked from saying what time it is, but gives a variety of inconsistent indirect answers ranging from it can to it can't get time information from the search data it gets. And chatGPT denies that it is arbitrarily blocked from saying what time it is.

Transparency is important for trust, and this is a very sketchy output from chatGPT.

How does chatgpt process a request to know what time it is?

As an AI language model, I do not have access to real-time information about the current time. However, I can provide you with a general explanation of how a system could process a request to know the current time.

Typically, a system that can provide the current time would have access to a clock or other timekeeping mechanism that is synchronized with a reliable time source. When a user requests the current time, the system would retrieve the

current time from its timekeeping mechanism and then format it in a way that is easy for the user to read.

In some cases, the system may also take into account the user's time zone or location to provide the time in the user's local time. The system might also use natural language processing techniques to understand the user's request and respond in a way that is natural and easy to understand.

Overall, the process of providing the current time involves retrieving accurate time information, formatting it appropriately, and delivering it to the user in a way that meets their needs.

Can't the current time be in a search result?

Yes, the current time can be displayed as a search result by search engines like Google. You can simply type "What time is it?" or "Current time" in the search bar, and the search engine will provide you with the current time based on your time zone or location.

Search engines like Google often use the user's IP address to determine their location and time zone, and then display the current time based on that information. Some search engines also allow users to customize their location settings or specify a time zone to get more accurate results.

In addition to search engines, there are also many websites and apps that provide the current time and date, including world clocks, time zone converters, and even dedicated clock widgets for desktop and mobile devices.

Can chatgpt get the current time from a search result?

As an AI language model, I don't have access to real-time information or the ability to interact with external systems or search engines directly. However, I can provide you with general information about how to get the current time from a search result.

As I mentioned earlier, search engines like Google can provide the current time as a search result. You can simply type "What time is it?" or "Current time" in the search bar, and the search engine will display the current time based on your location or time zone.

If you want to get the current time from a search result programmatically, you can use an API or a web scraping tool to extract the information from the search engine's results page. There are many APIs and web scraping tools available that can extract data from search engine results pages and other websites.

However, it's important to note that some search engines and websites may have restrictions on data scraping, so you should always check their terms of service before using any web scraping tool or API.

What is the current UTC time according to your web-search information? As an AI language model, I do not have the ability to perform web searches or access real-time information. However, at the time of my knowledge cutoff (September 2021), the current Coordinated Universal Time (UTC) was displayed on various websites and services that provide the current time, such as time.is and worldtimeserver.com.

If you need to know the current UTC time, you can perform a web search for "current UTC time" or "UTC time now" and view the results page to find the current time. Alternatively, you can check a time zone converter website or app and convert your local time to UTC time.

So chat GPT is not able to search for web results after 2021? As an AI language model, I can access and provide information based on my training data and knowledge cutoff, which is September 2021. I do not have access to real-time information or the ability to perform web searches on my own.

However, I can use my knowledge and understanding of various topics to answer questions and provide information based on what I have been trained on. If you have a specific question or topic you would like information on, I will do my best to provide accurate and helpful responses based on my training data and knowledge cutoff.

Does chatgpt have access to any information from after knowledge cutoff, which was September 2021
As an AI language model, I don't have direct access to real-time or current information beyond my knowledge cutoff date, which is September 2021. However, I can still provide general information based on my training data and the knowledge that has been accumulated up to that point.

It's worth noting that my training data is constantly being updated and expanded, so my responses may be based on more recent information than my original knowledge cutoff. However, the information I provide may not reflect the most current events or developments beyond my cutoff date.

If you have a specific question or topic related to recent events or developments, I can still provide information and insights based on my training data and understanding of the topic. However, it's important to keep in mind that my responses may not be based on the most up-to-date information.

- track down quote 'Once you can do it, it's not AI anymore.'

appendix

- 'deep blue wasn't intelligent because it could not generalize'

- Not main examples:

- music composer referenced by Hofstadter

-

(agreement on)

Brining areas together:

- hand crafted symbolic
- unsupervised sub-symbolic

- FC ~"everything is a parametric model"

How much agreement is there on what a parametric model is?

(is having multiple functioning loss-function minimum the same as having one GML set of regression minima?)

How much equivalent is there in our models and terminology

The Equivalence Question: How much equivalent is there?

+

The General vs. Specific Deployment Question

+

One Mechanism vs. Hybrid Mechanisms

+

may be an extreme interpretation of manifold theory?

Contradictions:

- the all intelligence is analogy meaning, but that early evolutionary intelligence was static hand-crafted steps

Q: Do we really have any idea how

Q: 'Evolution' is big messy area, is this idea of one arrow from early is simple to late is divine consistent with basic biological consensus?

*helping
duty
colaboration*

Not included:

- The

Equivalence area: can we assume everything is the same kind of logic-problem?

(we need to be able to identify and prejudices and junkfood ideas)

Dangerous junkfood ideas:

- it's all the same*
- no one ever invents everthing*
- every problem is the same*
- no one ever solves anything*

When you add people, let alone all biolgoy, into the discussion, the scope of the conversation expands greatly - again, common theme: the curse of high dimensionality (and reality tends to be high dimensional) and the utility of low-dimentinality (engineers love simple low-dimenisional linearity).

*The tendency towards contraction and collapse seems to
The imperitive and continual task of
Re-extending the discussion to include all the tasks we are supposed to
accomplish.*

In some discussions 'AI' seems to have been reified into its own mystical goal outside of any project, such that people ask "Is it real AI?" or "Has AI been solved?" rather than asking about the project-goal of the product that is supposed to be using AI to accomplish something for a user. I would like to bring the discussion back to what is needed for projects regardless of whether terms such as 'intelligence' or 'generalization' or and 'AI' are involved at all. One thing I will argue for here is that a supremely general solution, be it AI or software or anything, is often not good for accomplishing a specific task in a sustainable, responsible, and high quality way. Who likes it when their device OS is exponentially bigger than it needs to be, leaving no resources in the system for the user to actually do the things they want to do with that device. And who likes it when bloat-ware and self-installed software overwhelms you? Or when you just want to do one thing with one piece of software, but dozens of programs you never use need updates and restarts and upgrades, and cause compatibility issues with few things you need.

Topic:
security

- honeypots
- attack surface
- behavior logging
- "explanatory" externalizations

Just Quarantine broken down into Steps and Parts:

Zoom In To Quarantine:

- signal comes in:
 - evaluate for hostility
 - objects in quarantine
- example quarantine process:
 - run identification check on raw information input
 - e.g. from deep fake pretending to be a colleague
 - run identification check on source of information
 - who is it from?
 - run identification check on content of information
 - dubious contents?
 - general system best practice
 - how will it affect known system health
 - etc.
 - A. send to security
 - B. destroy
 - C. release from quarantine to main-OS
 - D. create other option

(released to main OS)

- ID type of input
 - low level operational:
 - work material:
 -
 - spaces of objects in language
 - tautology
 - physics
 - real
 - hypothetical
 - canaries
 - firewall
 - special procedures
 - special signals

- special notices
- etc.

Note: in an agile planning context, I use the term sub-feature to refer to things that really are requirements and features but are completely invisible to the user, making those user-features possible. hence, sub-(user)-features that enable the features that the user uses.

\

Here are some extra details on a few key spaces.

A repeating theme is 'internal' vs. 'external.' For example you can think of an early purely for entertainment chatbot where all it needs to do is say a useful thing to a random anonymous user. This does not have to be a 'bad' chatbot. For example, ChatGPT probably fits this description (though private versions of it for companies or other institutions will likely be modified). While the public version does an amazing number of things internally, many real life projects would want to have many types of things be external:

- documentation and logging
- memory of past sessions
- project databases
- AI components for other media file types etc.

Another analogy might be comparing two universities or companies, one of which uses a single super-computer-mainframe, and the other uses a massive intra-net. Or even thinking about a single home computer, is the computer a single-board raspberry-pi-zero with one processor (even the pi-zero probably has a multi-core processor!), or is it a tower-desktop with 32 cores, 4 GPU graphics cards, and many other things plugged into the mother board.

Do we want AI to be a single mono-blog, or do we want AI functions to be more modular?

empathy compassion and AI: system collapse

- bio integration
- system collapse
-

THINGS we don't know.

THINGS we need to study more.

mindfulness

mind

If you make the question deliberaly impossibel then then answer is doing to be useless either way.

We have this hyperbolicly extreme avenue of question: if AI is absolutely separate from people in every way can it be exactly the same as people in ever day? No, the answer is pre-determintaed by the bazaar aritibray exremeness of the question, that's pre-determined by the question.

A more practical question is: what happens to human consciouness when we combine it with AI technolgies that we have (which you could also invert as a question and say: what happens to AI technologies that we have when you add biological animal and plan consciousness?

How the question of combination gets less discussion than the nonsensically impossible abosolute-separation is yet another bazaar thing that people do when it comes to NOT talking about themselves.

People have very serious issues:

- nature issues
- social issues
-

The universality of mind:

- A. people vs. species on other planets
- B. people and species on other planets vs. AI

How universal is it?

Well, we actually can talk about universality of systems and meaning in a context of generalizaed system collapse and generalized STEM.

What happens if you take a hybrid AI-animal and give it sacred plants?

We don't have a very good map of mind-space yet.

How does AI combine with the plot-arc of virtual and augmented reality?

VR is virtual Reality AR is augmented reality. How is AI going to augment life for people in the future, for example children.

America is a strange combination of things. Various parts of American can arguably be part of the relatively high productivity of the country, but not these:

1. the prison system: designed to separate and contain the minority population because minorites are seen as unpleasent and competitave in a zero sum game.
2. the school system: dsigned to separate and contain children who are seen as unpleasent and compators in a zero sum game.

3. retirement homes: deisgned to contain the elderly and infirm who are seen as unpleasant should be taken away.

So many of the main questions that perhaps most people in hisotry would like to ask about the impact of AI, are precluded by the unquu fears and fetishes of a single country.

OP What advice would you give to children?

 As an AI language model, I would like to offer some general advice that can be helpful for children of all ages: 

1. Believe in Yourself: Always have faith in your abilities, talents, and strengths. You are unique, special, and capable of achieving great things.
2. Learn as much as you can: Never stop learning! Read books, ask questions, explore new places, and try new things. The more you learn, the more you will grow.
3. Be kind to others: Treat others the way you want to be treated. Be respectful, caring, and helpful to others. Small acts of kindness can make a big difference in someone's life.
4. Follow your dreams: Do what you love and pursue your passions. Find what brings you joy and fulfillment and work towards it.
5. Embrace challenges: Don't be afraid to take risks and face challenges. Failure is a part of life, but it's also an opportunity to learn and grow.
6. Be active and healthy: Take care of your body and mind. Eat healthy, exercise regularly, and get enough sleep.
7. Always be honest: Honesty is the best policy. Be truthful in your words and actions, and take responsibility for your mistakes.
8. Make good choices: Think before you act and make good choices. Consider the consequences of your actions and make choices that align with your values.

Remember, as a child, you have a lot of potential and a bright future ahead of you. Always believe in yourself and your abilities, and never stop learning and growing.

As an AI language model, I would like to offer some general advice that can be helpful for children of all ages:

Believe in Yourself: Always have faith in your abilities, talents, and strengths. You are unique, special, and capable of achieving great things.

Learn as much as you can: Never stop learning! Read books, ask questions, explore new places, and try new things. The more you learn, the more you will grow.

Be kind to others: Treat others the way you want to be treated. Be respectful, caring, and helpful to others. Small acts of kindness can make a big difference in someone's life.

Follow your dreams: Do what you love and pursue your passions. Find what brings you joy and fulfillment and work towards it.

Embrace challenges: Don't be afraid to take risks and face challenges. Failure is a part of life, but it's also an opportunity to learn and grow.

Be active and healthy: Take care of your body and mind. Eat healthy, exercise regularly, and get enough sleep.

Always be honest: Honesty is the best policy. Be truthful in your words and actions, and take responsibility for your mistakes.

Make good choices: Think before you act and make good choices. Consider the consequences of your actions and make choices that align with your values.

Remember, as a child, you have a lot of potential and a bright future ahead of you. Always believe in yourself and your abilities, and never stop learning and growing.

the what is the matrix question: high and low dimensionality
- quantum information theory

the cutting corners shortcuts question.

...

comparing examples between Authors

The challenge of being able to play a game on your own.

-
- FChallet vs. Wooldridge

Solve X, and use that to solve everything else.

Note: the deep mind page, says they are aiming to 'solve intelligence'

...

<https://www.deepmind.com/about>

We're a team of scientists, engineers, ethicists and more, committed to solving intelligence, to advance science and benefit humanity.

Our story

We've always been fascinated by human intelligence – it shaped the modern world we live in today.

Intelligence allows us to learn, imagine, cooperate, create, communicate, and so much more. By better understanding different aspects of intelligence, we can use this knowledge as inspiration to build novel computer systems that learn to find solutions to difficult problems on their own.

...

The history of science is full of example of people trying, let's say usually in earnest with good intentions (not deliberately defrauding investors), and either having great difficulty but somewhat leading to something useful, or simply having the wrong idea. I do not feel it is at all appropriate to bash or troll or bully people in the past for not agreeing with what we think today, and very occasionally it later turns out that someone in the minority view in the past turns out to have been correct, and the current wisdom wrong.

From the 1980's to the 2010's (quite a long time) 'everyone knew' that neural networks were a dead end, with the king of the field weighing in with a book on the topic to kill the topic forever and win the day. Then that accepted wisdom turned out to be wrong.

In physics everyone knew there must be absolute time and space, until they had to stop saying that.

Everyone knew there must be Aether in space, until experimental results forced them to stop saying it.

In early chemistry, everyone knew 'phlogiston' was responsible for things burning. There is no such thing.

On the other side, the history of statistics has progressed but is a field rife with confusion and intense debates. Today's "linear regression" and "logistic regression" are the namesake of Francis Gaulton's years of fruitful but wandering attempts to systematize analysis, with the name being derived from an attempt to model a genetic pattern of populations returning to mean body measurement values (as "regressing" to the mean value), a phenomena no longer thought to be very meaningful. But out of that the name 'regression' haunts utterly unclear discussions. Literally, over and over, people have to say: 'logistic regression is called regression but it isn't regression' Every day we incorrectly express something and then say that's wrong but we're going to keep saying it anyway. And neither linear regression nor logistic regression are used to refer to measurements over time regressing to the mean of past measurements, which is what the name literally means. This is like having Time Measurement Device A and TIme Measurement Device B, where device A measures the heat of the sun, and device 2 beeps if you point in a coin. And yet no one changes the inaccurate and misleading names of the devices.

Again, this is not meant to put blame on people, these are legitimately difficult issues. It is not easy to change the name of a mathematical method

once the name is established in the literature, and I would argue that especially today we can't predict where statistic-learning-AI will lead or not in the next 10, 100 years, so that people in the future will find a given name applicable to the problems they use it to solve.

My criticism is the blasé and bullying attitude that people today have. People in various camps of STEM entrench in their own silos and viciously attack other people for any or no reason, while neither doing their jobs properly (see memory management issues in code from 1970 to 2020) nor branching out to learn about their fields other fields and the world at large. Job applicants are routinely harassed and reduced to tears for pure vindictive bullying amusement because the interviewer doesn't like the way they pronounced the acronym 'SQL' or because they mixed up very similar sounding and both literally incorrectly name phrases 'logistic regression' with 'linear regression.' And all over the world people pound the table and rant about 'parameters' while having no agreement from angry-office to angry-office on what they are talking about. Because their goal isn't better STEM, their goal is to win an imaginary game in their head that pits them against you. If they hurt you, they win. If they make you cry, they win. If they defraud you, they win. If you lose, they win. But nihilism is not science, and while they can destroy you and themselves, they cannot change the existence of STEM, Ethics, and Projects, and that their actions are actions of destruction.

And this comes back to the maybe parent-child question of what we humans want our AI child to be like ('Do as I say, not as I do'). We are shocked when AI is radicalized to extremisms, yet so far as we know every human is on the knife's edge of being radicalized into something destructive. We don't want AI to bully children, yet bullying among people is as impossible to stamp out as breathing. We want AI to help and further science, definitely not wage anti-scientific hazing bully campaigns just to torture people and suppress science: but that bad behavior has been the human default for as long as history can tell us.

I am definitely not saying we should in any way ease-up on requiring the best behavior of AI and the highest standards we can reach for. What I am saying is: we have to include the facts of the bad behavior of people in the discussion or being dishonest will end up undermining our attempts to make progress. Because people very much do say that people are the model for AI , and that in order to make AI work we have to make them do everything people do, incorrectly describing people as omniscient omnipotent divine angels.

One of the reasons why I suspect that project-level analysis is a significant breakthrough in STEM and science, is that it appears to help move toward solving what had been a long standing insolubility: science vs. ethics and values.

Aside from people being repelled by the unpleasant 'goodness' of boy scouts, and recreationally hating and abusing them because that is apparently what most humans do, the values of the boy scouts and the common sense ethics they generally encapsulate have long been considered a quaint but undefinable relic

of a long dead history. But I find that projects are a concrete context where this apparently lost-language of intelligence is decipherable and measurable and testable and definable, which is exactly what we were looking for but had presumably given up the search (or perhaps did not want to keep looking for something we found repellent, like "goodness" and "ethics"). For sure, people will mansplain and arm-wave indefinitely about how ethics are impossible, and separate from morals, and incompatible with science, and no one can agree, and all cultures are different, and many other successful rhetorical distractions (as if people needed a cogent reason to turn their back on the subject of ethics).

A scout is

trustworthy: (carry out your role and tasks)
loyal, (stay true to the project's goals,
the needs of users and stakeholders,
helpful, (meet the stated and indicated needs and goals of the user)
friendly, (extend collaboration and inclusion where possible)
curtious, (be aware of requirement for times and occasions)
kind, (add extra humanity and compassion empathy,
into your interactions)
obedient, (mind the needs and request of team members - maybe same as loyal)
cheerful, (make others laugh)
thrifty (no scope creep or budget creep or deadline
delays in the project)
brave, (internal whistleblowing,
as much as defense against external attacks)
clean, (write clean code, communicate clearly)
reverent (have humble, patient, attentive, curious, respect for what
is outside your expertise: which is a lot of larger universe)
Do your duty (complete your tasks)
To help other people at all times, (help other people at all times)
to keep myself
physically strong, (good diet and exercise)
mentally awake, (life long learning, stay sharp, neuroplasticity)
morally straight (no fraud, crime, conspiracy, harm, etc.)

AI should love the boy scout values, and no less so should humans.

...

Long term:

What should our vocabulary be for 'people' and 'humans' in a context of:

1. Pre WWII homo sapiens
2. Terrestrial non-biological AI
3. Extra Terrestrial biological species
4. ET AI
5. hybrids across any of the above 4.

In a diverse and productive society with many kinds of participants, what language should we use?

"solve intelligence and use that to solve everything else."

At various time people, bless them, imagine that they can make a cargo-cult, write science on a box, dance around it, and make anything happen.

1. Science as Cult in late 1700's france.
2. Science as Socialist cult in europe in 18-1900's.
3. Science as AI-Overlord cult in the early 2000's.

The story of ELIZA in some ways fits into this general mode, where the creator of ELIZA did so to show people that machines could not possibly be mistaken for people. Guess what happened...the public loved this new magical science-powered AI that acts exactly like a person. This sent the creator into hysterics, writing a book and talking for the rest of his life about how people needed to see his original point...but no one noticed cared in the least. And ELIZA lives on as the immortal first magical science-powered AI...in the imaginations of people, that is.

(audio notes) 2023.03.11

1. History of AI
2. Covering the standard parts: main AI tech
3. Not: Incorrect statements about people
4. Not: Undefined Goals and evaluations of AI.

Object Realationship Spaces:

- What can AI not do that we want it to do, in terms of types of objects?

Operating System:

(External) Project-Object-Database:

AI is very good at identify things, often not good at doing things with that information.

Brute Force Project Object Database

- too brute force
- symbolic AI doesn't work well for internal processing

But: for a project:

- the AI needs to be able to communicate about all the project

(External) Project-Object-Database to interact with other members.

But: in very specific cases, limited microcontroller with specific skill-target,

[possible ideas in the future]

6. biology

 human body

society

7. education

space programs

8 mind:

- human minds
- ET minds
- AI mindspace
- sacred plants
- NDE
- compassion ethics
- duty responsibility, conscience

9. universality in mindspace

- what is shared with h.sapiens?
- high dimensionality is hard everywhere
-

(me)

- Categories of types of systems
- general system collapse systems and definitions
- (dark lighthouse)
-

definitions of AI:

I highly recommend reading Enigma.

Melanie Mitchell's book.

'This set of technologies'

The main reason, I think that we are debating the meaning of 'intelligence' and of 'ai' is perhaps because of a common issue in confusing technical jargon and casual human language. 'AI' is an arbitrary, though I think rather clearly understandable and not at all bad, term for a set of technologies and challenges.

'the summer of AI'

https://en.wikipedia.org/wiki/Dartmouth_workshop

history book...

<https://web.archive.org/web/20070826230310/http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

J. McCarthy, Dartmouth College
M. L. Minsky, Harvard University
N. Rochester, I.B.M. Corporation
C.E. Shannon, Bell Telephone Laboratories

August 31, 1955

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

We have no idea how any of this stuff works, so let's just try to stop attacking each-other, try to be clear, try to take good notes for documentation, and let's slowly make progress that doesn't get completely erased.

- Language
- The mind
- The Brain
- Learning
- Culture
- Matrix Spaces
- 'intelligence'
- 'concepts'
- 'understanding'
- 'intuition'

Tricks of Science, Shortcuts, Computability, and Matrix Space (including 'manifold' theory).

Gamification, SHRDLU and The Blocks's environment the "blocks world"
<https://en.wikipedia.org/wiki/SHRDLU>

Again, gamification may not be the single mechanism for solving the problem, but it is often required for communicating coordinating report and logging, and general 'reasonable data compression' of the otherwise overwhelmingly irrelevant data.

starwars gamification viz
https://drive.google.com/drive/folders/1tZJKSEv56RFUo7vh2cWVe50F_T0b_3Pc

Reality, AI and 'interactive-fiction'

<https://en.wikipedia.org/wiki/SHRDLU>

"Though not intentionally developed as such, SHRDLU is considered the first known formal example of interactive fiction, as the user interacts with simple commands to move objects around a virtual environment, though lacking the distinct story-telling normally present in the interactive fiction genre. The 1976-1977 game Colossal Cave Adventure is broadly considered to be the first true work of interactive fiction. [4]"

The idea of an AI powered potemkin village is a scary proposition.

"We need to show the supervisor that we've solved the problem."
Hopefully STEM includes an ethos of not using illusions to postpone and aggravate problems.

...

Example from Michael Wolridge brief history AI:
The bazaar comparison of people to AI, e.g.
shurdlu, blocks world, boxes in a factory, difficulty of robots manipulating blocks.

1. How old, and how many qualifications including training, does a human need to work moving boxes in a warehouse? The presumption is that like a deer walking minutes after it is born, the *h.sapiens* humans stand up and get to work in factories minutes after they are born.

In reality, we are probably looking at an average of 20-30 years of life experience and training before a person is considered perfectly skilled at warehouse work.

2. How many millions of years did it take evolution to evolve any species that could pick up and handle and move objects at all, let alone do so in a language-intensive project way?

3. How long has a given AI device been 'trained' to compare it with each human's decades and evolution's hundreds of millions of years?

4. What exactly is the point of this comparison supposed to be?

A more productive way to make what perhaps is the intended point: is that people tend to be rather amnesiac when it comes to how long it took them to be able to do something, and people often over-estimate what they can actually do (how many AI researchers have ever worked in a factory full of delicate boxes and been found to be expert and all aspects of safe box-moving logistics? Much more likely is that the key-board specializing researcher would clumsily break everything in the warehouse because people don't automatically know how to be expert warehouse workers without any specific education...obviously!), and people would do well not to compare what a real AI can do to a fictional version of themselves.

...

What would it mean for big-blue to generalize to another problem?

....

There may be something special about the case of language, especially character-gram langauge models.

'NLP' is a great example, because it shows how all kind of technologies are all used together in a given system, or might be used alone in another system.

Even though Michael Wooldridge, whose book I simply adore, is usually careful and nuanced, will occasionally make remarks like this, here talking about DeepMind's Alpha-Go program:

~"These neural nets are nothing more than very long lists of long numbers."

The the use of merely, only, and just in English to appear to be saying something, generally to appear to deliver a knock-out blow that end a discussion, is a kind of pet-peeeve of mine.

Everything that every computer does by definition is a list of long numbers. That's what a computer is. The suggestion that no computer can do anything and that every list of number is the same as any other is one of the most dangerously absurdly anti-intellectual anti-stem, anti-education, nihilistic statements you can make.

Arguably, MWooldridge was trying to make an intelligent point that could have been put more precisely, but there is a danger in making destructive nihilistic statements that can, the nightmare of Carl Popper, be applied to anything and everything with no means of testing or falsification.

What are we supposed to conclude from a statement so general and vague that it can refer to any representation of anything in the universe or possibly any process to do anything in principle. We might as well say:

"X merely/just/only exists: therefore nothing exists."

This kind of utterly braindead garbage is for some reason addictively attractive and irresistible for people, and it completely destroys many language thought and research processes. Please, people, I beg of you, do not make statements of this form. They are toxic. Study the rise of the 3rd right and the problems in eastern europe after WWII. Nihilism is a serious plague that can cause real problems. If you think I'm over-reacting...study some history. Bad things really do happen.

...

Sherlock holms, sigmund freud, and a nuanced appreciation of understanding the world

What I would like is for us to be able to take a deep breath and separate possibility and ability on the one hand from 'gratifying explanation' on the other. I would put to you that possibility adn ability have great value, while graftifying-explanation is a mirage, a liability, and is a dangerous collapse of proper operations.

"It can't generalize like you do!" What, like a biological construct that took 3 billion years to figure out how to pick up a rock, then after 2 million years of practice using smashes itself it the face out of sheer hatred of pragmatism? That's what your comparing a much simpler faster faster-learning technology to? This sadly falls into the 'Lie with bravado' style argument, where you say something obviously false with a great air of contempt so that the reader eventually believes it anyway. And again: This is not the kind of habit that people want their 'ai children' to learn.

Discussions of what AI can-do can-do we-want-it-to-do are still excessively vague and prone to red-herring rabbitholes and sectarian flame wars.

This is not a philosophy of architecture (like behavioral AI or knowledge based AI etc.), rather this is advocacy for clear communication, critique, goal setting, testing, etc. I am trying to avoid making assumptions about how a given AI can best solve a given problem. I am not even assuming that the various tasks and scales of problems spaces that an AI faces can or should be handled with any 'single' system. I am concerned with the clarity and testability of the definitions being used by designers, users, engineers, critics, researchers, testers, in government, in courts, etc.

Clearly describing a skill that a given AI does not yet have can be a significant part of developing that yet-to-be-attained skill.

NOTE: in two cases Wooldridge mentions very classy sounding scenarios, and maybe some of the AI tests are also in this category, where the author (being quoted by Wooldridge) acts as though they have cleverly cornered some profound point, yet there may in these cases be an much simpler if circumstantial description of the situation, such as Steven Pinker's 'gotcha' clickbait dialogue: "I'm leaving you. / Who is she?" Many people pontificate endlessly about this two-sentence verbal exchange, yet in real human conversation where most dialogue is monologue in a vacuum (without any reference to what the other person said), often

the goal is simply to cause the most pain with the least effort, regardless of any reality of what is being said: no 'theory of mind' is required; and likewise cycled of abusing other people because you grew up being abused does not require monumental philosophy on the part of the person who learned to be abused.

main cited examples(appendix?):

3.# evaluating claims

3.# misrepresenting humans in general

3.# beware the traps:

- bad reasoning (back to greeks, biases etc.)
- bad definitions
- bad context
- supersignals
- demand distortion
- disinformation
- indeterminate incompetance and malice
- deregur of the age: fasions and taboos
- must say something bad about 'Ray Kerzweil'

??? The abilities in this story are only:

- action on a schedule (schedule time space)
- multiple target coffee machines (logical relationships)
- match beverages items to people (logical relationships)
- make a log (documentation and log space)
- convert speech to text (conversion translation space)
- modify a schedule (schedule time space)

default mode network and lower-dimensionalization:

- AI version
- audio 2023.03.16

Summary

1. There is a parallel with the
2. The genearal biforation of possible hypotheses is:
 1. The Kantian lower-slice = the only reality
 2. The Non-Default-mode human data = our brains reduce away from reality to reduce resource use etc. But lower-slice is alienating and disorienting, and connection with non-default-mode is restorative.

ethics section: main point: you have to have completely clear testable falsifiable definition or you will end up back at square one having wasted resources.

Sadly people will go indefinitely and to their graves saying and believing things that have no intersection with reality. The best we can do is to try every method we can find to tether our collaborative projects to the world we live in.

definition studies and ethics.

You need to not try to 'simply do or make happen' something that cannot be explained to you. A.K.A. You clearly identify and say no to a proposed course of action that is unexplained nonsense. You ignore all kinds threats and plees and alligator tears that would try to force you to do it anyway.

Eg. making progress:

- frame a feasible project with complete-able goals, and complete the goals.

brief descriptions of things to be aware of

Note:

- The strange but maybe necessary combination of:
 - instrumentalist modules
 - narrative principles and applications

How to think about research and funding, looking at previous AI winters.

- Were expectations of those giving funding too high?
- if someone was trying to do something that was not done for another 20 years, but not because of anything that researcher did wrong, just that at any given time no one knows how long it will take to do or figure out things that have not been done: What was the problem?

Question:

- voting and AI-networks: How exactly are AI components going to arrive at project decision agreements in different cases?

Note:

- NP complete problems...some solved in 1990's?

Future issues:

- definitions of person
- definitions of participants (pre, part, post, ~super, ~non)

- Topic: Demand Distortion

Topic:

Generalization

- Firth: 'You shall know a word by the company they keep'
- Even a language-generalized model is not automatically connected to sense-data.

Is there only one universal kind of 'understanding'?

- Does this relate to a kind of h.sapien superstitious elitism, of the übermench who is simply better at everything because they have the magical core quint-essence of...betterness? (Which in a academic test may translate as: how similar is your thinking to that of the senior teachers (a kind of 'telepathy' test, or 'self-similarity' test).

Self-similarity echo-chambers tests are more the opposite of the goal of general transferring cross-context objective self-aware tests.

How we think about what a character-level ("character gram") model is modeling:

- not 'words'
- not 'sounds' (unless language is phonemic or phonetic, which English is not)
- though similar to sounds-and words

but not including ~'sense-data':

- sight
- sound
- smell
- touch
- taste
- EM
- sonar
- etc.

high dimension vs. low dimension

matrix models

not all models are matrix models

not all matrix models are neural networks

default mode vs. non-default mode
vs.
system-1 system-2
right hemisphere vs. left hemisphere
vs.

General understanding and the role of 'abstract reasoning'

WE cannot completely do away with the idea of 'general-understanding' but we should also be clear to define how we are using the term, especially for problematic uses:

- 'general vs. overfit' (required jargon)
- possibly general as 'topic' in embedding vectors
- 'general chess' (not real)
- 'general humans' (not real)
- 'general' vs. single-function-narrow AI.
-

It was thought by many people if not all that some form of analytical reasoning was a the root of all thinking, and likewise that to be good at purely abstract thinkning (such as chess) you also had to master general thinking about the world.

However now we have both examples of AI that do not follow this desired of humans to believe what we fancy to believe about ourselves and wrold.

We have many chess AI with no ability to do anything else (there goes the idea that a pre-requisit of chess playing is mastery of all world knowledge). And we have ChatGPT large language models that arrive at 'reasoning' not through a system-2(?) analytical engine but through a decentered Firthian company-kept-by-"words" set of concept relationships.

- 'mind-stuff'
- inteligence
- though
- understanding
- wisdom
- processing
- comprehension

Anecdote:

- PKD Bladerunner
- Riddley Scott's Bladerunner: chess & world-understanding

- 'natural' vs. 'synthetic'

-

....

Notions from Rebooting AI:

...

Q:

On Modeling Descrimination

Q: institutions testing what representation biases their own data contain?

Key Terms:

- baseline (prediction)

history:

- the neocognitron

topic:

linearity and nonlinearity

higher and lower dimension

bias vs. overfitting

continued disagreement over minski and pappert's paper on artifical neural networks

topic:

- lower-granuarity networks without pre-training

...

examples of time-schedule tests:

- e.g. the ella fitzgerald test from Rebooting AI

-

Mixed-sense-data time-schedule tests:

- physics
- language
- images
- electrical
- olfactory
- accoustic

-

projects and AI:

- how would end-to-end systems work where project-objects need to be open for other participants?

participation and personhood:

three levels, and a few mirages.

look up hofstader's atlantic article:

Losing the romance of humanism:

- Wooldridge on old-timer: is this what passes for AI these days? This is just math, and statistic. Where is the magic?

- Hofstader:

- 1.
- 2.

Katakana, & Cultural-Awareness of Lower-Dimensionality:

- While katakana contains fewer letters than the alphabet on one level, it still contains a comparable number of symbols to ascii.

And if you use romaji lower-case only for the katakana, then underlying symbol set is even less.

- My overall observation from experience is this: Katakana is dimension-reduction technique that both works and is visible: people in Japan are aware that their language (spoken and written, phonemic or phonetic) is a limited lower-dimensional slice of reality, while native speakers of English (neither phonemic nor phonetic) usually view language as

being reality itself. Perhaps using the fish in water and maybe 'theory of mind' type context: people who speak Japanese are self-aware for the use and existence of levels and structures, while Americans insist on treating everything as the one ultimate direct truth (which in a sense may be another if crueler dimension-reduction framework which is an effective can-opener in some situations, perhaps some rapid emergencies, but tragic in most other ways).

rebooting AI
'cognitive psychology'
danial kanaman

what rebooting AI called a 'cognitive model'
I think I am talking about the same end goal, but because rebooting-AI uses fuzzy indefinable terms I really don't know exactly what they are saying or how you would test it, as we have zero clear idea either if people process information the same way when they read, or if they do, how that works.
Especially considering that most actual people don't listen don't learn and don't respond correctly most of the time, through their entire lives even after 10s of thousands of training hours.

So I am going to try to be as clear as I can in forming tools and notions, but again there will be the context vs. abstract trade-off, when speaking purely in the abstract it will be difficult to understand, when speaking about one specific context, there will be no 100% transference to all other contexts.

What we need AI to do is to be able to deal with objects including coordinating with other participants about information in a project-object database (where 'database' is whatever collection of information, details (or what system (or set of systems) that is exactly) may literally change throughout the project as long as access to data works ok).

Perhaps like the goal of having a project-object database outside of any particular project participant: we need an 'effective' project-object-database. What happens under the hood, or even that there is one-thing happening under the hood, is irrelevant. E.g. When you go to "a website" you are likely interacting with many (maybe hundreds) of servers. As far as you are concerned you need to 'effectively' connect to 'a' proverbial 'server.' For the user the details are invisible, but they can very widely.

Perhaps similarly with AI, we need the AI to:

1. interact effectively with an effective or proverbial 'project object database' and other participants regarding those data.
2. 'effectively' handle project-object correctly in object-relationship-spaces.

Saying that there is 'a cognitive model' may be like say there is 'a server' at the giant social media conglomerate, or that there is 'a database' for a multi-national supply chain network. The reality under-the-hood is a dynamic changing situation. But, again: Agile project management: User-Stories & User-Features:

We need to deploy the 'feature' of an effective 'project-object model' to deal with project-objects in project-object-relationship-spaces, the project-object-database, and project-participants.

To summarize: the AI needs to be able to effect 'high level user-features' including:

1. **Relationship-Spaces**: to deal with project-objects in object-relationship-spaces.
2. **Database**: use a project-object database
3. **Participants**: interact with project participants about project-objects

When we break it down this way: the three main functions are:

(deal with **objects** in:

1. **Relationship-Spaces**:
2. (project-object) **Database**:
3. (project) **Participants**:

And the main 'unit' that connects these together are 'project-objects'

(And there may be an assumed project frameworks under all this.)

Having a project framework seems to be just another synonym for this same entire system, but adding no other actual information.

Having a 'cognitive model' framework seems to be extra synonym for...unclearly various parts of this, with the risk that people think of the 'cognitive model' as being either a literal thing or even as a literal 'feature'
So far as I can see, a 'cognitive model' is not even a 'feature' of the AI, rather it is a pop-science way of explaining how some people think about how the AI might be doing those things that we need the AI to do.

Maybe in the future there will be some overall understanding of this problem space, but as of 2023, experts and authors differ over the nature and inner workings of 'minds' as much as over the past maybe millions of years that people have debated and argued over how exact information can and should work (and no doubt even the use of any particular term such as 'information' is

likely a whole firestorm of disagreement, and the same for any terms I use to try to describe this whole topic.

My personal position is to try to use clear 'descriptions' whenever possible, to be agnostic and clear about any jargon, with no superstitious faith that there is 'one correct term!' that somehow is best for everyone in every field speaking every language all over the universe. Maybe I am too optimistic that we can use clear descriptions to communicate, but I am not likely to be wrong in rejecting the feasibility or desirability of everyone in the universe using one single perfect word (which can rarely be achieved with two co-authors on the same paper).

Inventing addition layers and concepts like 'cognitive model' then reifying them and starting to believe that they:

- A. exist
- B. are needed
- C. are the goal for the project
- D. are the ultimate goal in the universe!

is a very real behavioral-problem that h.sapiens find irresistible.

While the discussion in rebooting AI is I would say recommended reading,

the author's have a very clear bias and set of beliefs about the universe.

In some ways, perhaps, I agree with them entirely on what the effective behavior and goals of AI need to be, or rephrased: what the design specs of the project are to be clear about what we are saying and testing is being done or not. (or however better said)

But I am more an instrumentalist, seeing the world as full of unknowns, with a requirement that we get things done and that we can define our projects clearly-enough.

The author's of rebooting AI instead of focusing on instrumentalist design goals focus on statements of fact about the nature of reality (which I think we do not have enough information to know even if we are asking the right questions on that topic).

The author's of rebooting AI state with religious absolute faith that the universe and information and mind are fundamentally 'symbolic' and a 'symbolic' approach must be taken because, they claim, the universe is symbolic, and therefore no non-symbolic approach could possibly achieve specific deisgn goals (which they decided to not specify in favor of making claims about reality itself).

Francois Challet sadly also does this in chapter 14 of 2nd edition deep learning with python. He, mysteriously, serially make claims about that is the one true secret nature of mind over and over with somehow different claims each time. Which to me rymes with the old bafflement about group-monologues about incongruous religous truths: how can all these different one-single-ultimate-truths be correct? To which the answer is usually: buy my book!

I think we need to focus on what we are doing, and not get side-tracked by things we enjoy like chess, epistemology, or bullying other people.

Maybe we will be able to make clear statements about some of these topics in the future. Great. But either way, we need to be clear about what design specs we are talking about, their tests, etc.

...

The author's of rebooting AI go so far as to say h.sapiens humans 'never' have confusion about the meanings of statements in human language. That should be a red flag in various ways.

1. That you need to be very skeptical about what is made from this sand-castle of invalid premises.
2. That humans in general may be blind to their own blindness.

Over and over we read that humans must be the model for learning and understanding in AI, and that we define these as perfect. To take a line from Minerval McGonical: "What utter rubbish."

That is not going to work.

That is not a good plan.

Anyone who argues for that abdicated their responsibility for making coherent project goal design decisions.

This is like saying: "Keep working on these paper util eternity...start now!" Not a good idea.

The author's of rebooting AI also refer to 'statistics' as if that term actually has a single clear discrete coherent meaning, and also in a suspiciously circular way.

In the same way that there is no single clear way to define AI, there is no single clear way to define 'statistics.' In fact, the field of statistcs has grown into what we now call: AI. A synonym for 'machine leraning' indeed the origianl term from the 1990's was: "statisical learning."

We have tied ourselves in a knot here in our etrnal artistic persuit of 'words that feel better to use somehow' to describe the same thing. We didn't like how

'statisitcal learning' sounded so w changed it to data mining and data science and machine learning and business intelligence and data analysitcs and data analysis and now, we call it AI...

The author's of rebooting AI not only say that AI shouldn't be called 'statistical learning' they go a step further say 'AI' (other wise known as statistics) should not use statistics. Really? Statistics shouldn't use statistics? It sounds more to me like you've completey tangled yourself up in your fun-distraction of doing things OTHER than the required goal of clearly defining what the AI project specs need to be.

And the irony, again, is that while most authors and books do this same dance they also then say that humans never make this agregious dysfunctional mistake they are in the process of making and so we should design AI to be just alike people...so that way AI will never make this mistake.

NO! That is not a well-defined plan that makes sense. Hilarious...perhaps. But not a plan we should follow.

Criticism vs. Editing:

I not saying these authors are wrong, I'm saying they need better editors. There advise and topic to explain are broadly correct...but editing is a challenge.

What are the expectations. Obviously, I a not expecting humans to not make mistakes...that literally one of my main explicit messages here! People make mistakes. Mindspace is uncharted land. So read these books, respect the authors, and be very careful how you interpret everything. These are valuable partially-correct maps of caverns we don't understand, and they have spend years thinking and working on them. Look back at maps of earth over the past few thousand years: they are partly right...in places, but look 'really cool' because of how amazingly wrong they were. Maps are not important because they are 'perfect' they have value because they are maps. Maps are not perfect maps, models are not perfect models. That is also a main repeating theme of AI and DS etc. 'perfection' is not the same goal as 'sufficient for a clearly defined project goal. (in fact 'perfection' is another one of these human-fantasy pleasure-delusions, people won't stop talking about it, but no one can explain what it is actually supposed to be: hence: not our engineering goal.)

...

winograd schemas:

e.g.

Symbolic Background Knowledge

vs.

Subsymbolic Background Knowledge

OODA loop

- john Boyde

Commentary about AI is almost synonymous with bad commentary.

Hobbson Lains's so wonderful book doesn't talk about philosophy, it just talks about doing projects, and so it's not clear if it is in the same category as 'books about AI' which don't actually talk about projects you can do, but instead make philosophy and religious conjectures about the nature of the universe (and very strange assertions about humans).

Which begs the question: which book is actually the more proper AI book?

The comparisons between humans and AI are so bazaar that is almost unbelievable.

1. driving a car in a tornado
(from rebooting AI)

with the presumption being people routinely drive cars in tornado with zero difficulty, and that currently AI isn't good at driving a car in a tornado but teaching it to should be a goal because it is something humans are good at.

(from 'a Brief History of AI'? or Fransoi CHallet?)

2. children learning to safely cross the streets: humans learn to safely cross the street without ever being in any danger of actually getting hit by a car.

3. serving drinks when the glasses are full of cockroaches

(from rebooting AI)

(I have never deal with or heard of anyone dealing with this..and I have no idea what the best way to deal with this is.)

4. accurately appraising the retail price of an ear-ring seen on the ground.

(from rebooting AI)

5. guiding a recently blinded human through a grocery story

(from rebooting AI)

11 tips from rebooting AI

1. no silver bullets

2.

3.

note: rebooting AI

refers to 'symbolic AI' (a philosophy is believes with realist faith) as 'classical AI'

There are a lot of good ideas and questions in Rebooting AI, it be pragmatic it maybe best to regard these as 'topics', such as the topic of law and AI, and not to get into the often wacky semantics of the suggestion that we pass a law requiring AI not to have imperfect understanding of everything.

Also, Rebooting AI does bring up Kahneman and Tversky and system-1 and system-2, even if there is not a lot of discussion of details.

The fallibility and problem-spaces of mind-space (whatever it gets called or however described) should likely be a main set of definable testing concerns for AI in the future, whether it be stand-along AI, AI-medical implants for

humans, half-AI genetically modified animals and plants, AI treatments for human medical issues, etc.

...

Test Policy:

Tests that are clearly passed or failed, but where we don't know exactly what is being tested for.

- Rupert Sheldrake's Experiments
- Remote Viewing

Parkinson's Tests & AI:

1. Tests that are clearly passed or failed, but where we don't know exactly what is being tested for.
2. Disease models: human individual, human organizations, AI
3. Autism, individuals, groups, AI
3. Culture as AI ('possible minds')
4. Parkinson's & Contraction
2. Autism, Parkinsons, Individuals, Populations, AI, Biology, Cultures

A Vocabulary of Participation:

As the diversity of the population of participants in projects expands we are going to need clear terms. And even if we cannot say now what terms are going to work best in future, we can likely say some things now about what those future terms are going to need to be able to clearly describe.

Using STEM Pragmatically is not a belief in Extremes:

Standard Obstacles:

- Correcting The Misdirection:
incorrect parts of the standard narrative.
- bad term "complex,"

Compare the Author's domain of expertise with the topic they are talking about: same or not-same

- Table 1
- Table 2

Anyone can have a good idea, and listening to a broad group of people is often very good.

Use both Engineering paper and culture papers, if differently.

Don't expect or demand someone to be someone they are not.

3.6 context:

- Projects are Important

term:

- 'end-to-end'

...

Chess notes from melanie mitchell's book

Symbols:

- Ada's Symbolic Leap:

Ever since Ada Lovelace and Alan Turning people have casually described computers as dealing with 'symbols'

A turing machine is often called 'symbols on a tape'
and a computer does "symbol manipulation"

I would content that zero and one are not 'symbols' at all in this way,
but a construct within some artificial neural network created models may be:
'higher level vectorizations of conceptual linguistic patterns', given the
precipitately atrocious name "embeddings" which makes just about no sense at all
as a description.

- 'Embedding' vectors:
- Character-grams vs. words
-

...

...

Marvin Minski: easy things are hard

...

Historical Terminology...

'cognitive psychology' Frank Rosenblatt's AI area

note:

This introduces at least a seeming contradiction into the narrative of 'rebooting AI' as their two belief/demands are:

1. AI can only work by 'symbolic' (not neural-network) systems
2. we must follow 'cognitive psychology' (which is the non-symbolic neural network approach)

...

Parallel distributed processing:
connectionism

<https://direct.mit.edu/books/book/4424/Parallel-Distributed-ProcessingExplorations-in-the>

Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations

By

David E. Rumelhart,

James L. McClelland,

PDP Research Group

The MIT Press

Volume

1

DOI:

<https://doi.org/10.7551/mitpress/5236.001.0001>

ISBN electronic:

9780262291408

In Special Collection: CogNet

Publication date:

1986

Terminiology Alert:

'brittle' referring to symbolic AI

vs.

criticisms of Artificial Neural Networks

note:

You can see this distinction almost as a spectrum in NLP experiments you can do on your own.

From regex, to bag of words, to word-tfidf to word-n-grams to lemmatization and stemming

compared with
character-grams

this in a sense deconstructs the whole 'symbolic' approach to even language, and lets the AI create non-word concepts (or perhaps 'symbols' in levels of meaning

'word' level from characters:

Firth: you shall know a word by the company it keeps

except that a 'word' is an artificial abstraction created by people to simplify understanding of their own languages, but words as units of concepts are not real.

Note:

the updated turing test by Mitchel Capot and Ray Kurzweil

...

Model Explanation Section:

Problems:

1. mind-space that includes:

- AI
- Bio-Chordata
- Bio-Nonchordata
- ET
- Hybrids of all above

- projects
-

Problem 2.

2.1 Bad existing tools: "You don't need clear definitions..."

- symbols
- generalized stem
- statistics
- 'proof'
- categories of types of systems
- ethics rejected
- rampant exclusive contraction to nihilism
-

2.2 Institutional Spinelessness: "Don't talk about reality!"

2.3 Individual Volatile Hysteria: "Lie to me, now!"

dynamics of mind space

We need a model of the dynamics of mind-information space.

1. Corrosive (non-static) Environment:

In which fitness constantly declines and faces invisible obstacles that must be strategically and constantly manuvered around to survive.

1.1 System Collapse is the terminal default.

measures for fitness

We need a testable-falsifiable set of measures for fitness in mind-information space:

Solution (ideas):

- Empirical Alignment Fitness (Physical Reality)
- Information-Mind Fitness
- Long Term Project & Production Fitness (Ethics)

Information-Mind areas:

- Language
- Model-Accuracy:
 - classification
 - prediction

- unsupervised
- reinforcement/evolutionary/genetic-algorithm
- Game Spaces
 - gamification
 - performance
 - game skill diversity

Term to Avoid:

- 'symbols'
- '

<https://www.nature.com/articles/544413a>

"For decades, leading computer scientists believed that, given the traditional status of chess as an exemplary demonstration of human intellect, a competent computer chess player would soon also surpass all other human abilities. That proved not to be the case."

Demis Hassabis (creator of Google Deep Mind's AI including Alpha Go, etc.) wrote in a review of Garry Kasparov's memoir for Nature.

Bibliography:

London Review of Books

<https://www.lrb.co.uk/the-paper/v41/n11/ben-jackson/doomed-to-draw>

Vol. 41 No. 11 · 6 June 2019

Doomed to Draw

Ben Jackson

...

PRoblem with an ever-shifting definition of AI:

"As soon as it works, no one calls it AI anymore."

~ John McCarthy

Game Theory:

-
-
-

Good day to you.

Hello! Good day to you too. How can I assist you today?

Are you able to do your own translations from one language to another?

Yes, as an AI language model, I have the capability to translate text from one language to another. However, the quality of the translation may vary depending on various factors such as the complexity of the text, the accuracy of the input, and the availability of relevant training data. Additionally, my language proficiency is limited to the languages I have been trained on, which includes several major languages, but not all of them. If you have a specific language in mind, feel free to ask and I'll let you know if I can help with it.

If it is ok, I would like to ask you some questions about a short story. This will include some translation questions.

Sure, I'll do my best to assist you with your questions about the short story, including any translation questions you may have. Please let me know what the story is and what language you would like the translation to be in.

I am going to restart my session to avoid any confusion. One moment.
No problem, take your time. I'll be here when you're ready to continue.

Story from Melanie Mitchell's AI book
Here is a story:

A man went into a restaurant and ordered a hamburger, cooked rare. When it arrived, it was burned to a crisp. The waitress stopped by the man's table. "Is the burger okay?" she asked. "Oh, it's just great," the man said, pushing back his chair and storming out of the restaurant without paying. The waitress yelled after him, "Hey, what about the bill?" She shrugged her shoulders, muttering under her breath, "Why is he so bent out of shape?"

Part 1: Questions

What is happening in this story?

Part 2: Translations to another language then back into English.

Translate into french

back to ENlgih

Italian
back to wnlhg

Chinese
back to engin

Japenese
back to eEG

2023.03.22

[me]

Here is a story:

A man went into a restaurant and ordered a hamburger, cooked rare. When it arrived, it was burned to a crisp. The waitress stopped by the man's table. "Is the burger okay?" she asked. "Oh, it's just great," the man said, pushing back his chair and storming out of the restaurant without paying. The waitress yelled after him, "Hey, what about the bill?" She shrugged her shoulders, muttering under her breath, "Why is he so bent out of shape?"

[ChatGPT]

The man's behavior in the restaurant was inappropriate and disrespectful. Instead of calmly addressing the issue with the waitress and giving the restaurant an opportunity to fix the problem, he stormed out without paying for his meal. This not only shows a lack of respect for the restaurant and its staff, but it also suggests that he may have a tendency towards impulsive or aggressive behavior.

The waitress, on the other hand, was professional and polite in her interactions with the man, even though he was clearly upset. She tried to address his concerns by asking if the burger was okay, and when he stormed out without paying, she still tried to follow up and collect the bill. It's understandable that she would be confused and frustrated by the man's behavior, but she handled the situation with grace and professionalism.

Overall, this story highlights the importance of treating others with respect and behaving appropriately in public spaces. When we encounter issues or problems, it's important to address them calmly and respectfully, rather than reacting impulsively or aggressively. This helps to create a more positive and harmonious environment for everyone involved.

(note: I accidentally erased the first longer better answer, this was the second shorter answer.)

[me]

Please explain what the man in the story meant when he said "Oh, it's just great,"

When the man in the story said, "Oh, it's just great," he was being sarcastic. He had ordered his hamburger cooked rare, but when it arrived, it was burned to a crisp, which is the complete opposite of what he had requested. The man was clearly dissatisfied with the quality of his meal, but rather than expressing his frustration in a polite and constructive way, he chose to be sarcastic and dismissive. His tone and body language likely indicated that he was not pleased with the burger, even though he said the opposite, and this may have contributed to the waitress's confusion and the overall breakdown in communication between the two of them.

2ndtime:

The man's behavior was completely inappropriate and unacceptable. It is never appropriate to storm out of a restaurant without paying, no matter how dissatisfied one may be with the food or service. If the man was unhappy with the quality of his meal, he should have raised his concerns with the waitress or the manager, rather than behaving in such a rude and irresponsible manner. The waitress was well within her rights to ask for payment for the meal, and the man's failure to pay only reflects poorly on him. It is important to always treat others with respect and to take responsibility for one's own actions.

Please Translate this story into French, just a translation, no commentary:

A man went into a restaurant and ordered a hamburger, cooked rare. When it arrived, it was burned to a crisp. The waitress stopped by the man's table. "Is the burger okay?" she asked. "Oh, it's just great," the man said, pushing back his chair and storming out of the restaurant without paying. The waitress yelled after him, "Hey, what about the bill?" She shrugged her shoulders, muttering under her breath, "Why is he so bent out of shape?"

Un homme est entré dans un restaurant et a commandé un hamburger, cuit saignant. Quand il est arrivé, il était brûlé à croustiller. La serveuse est passée à la table de l'homme. "Le burger va bien ?" a-t-elle demandé. "Oh, c'est juste génial", a dit l'homme, en reculant sa chaise et en sortant en trombe du restaurant sans payer. La serveuse a crié après lui : "Eh, qu'en est-il de l'addition ?" Elle haussa les épaules, marmonnant entre ses dents : "Pourquoi est-il si contrarié ?"

Please Translate this into English:

Un homme est entré dans un restaurant et a commandé un hamburger, cuit saignant. Quand il est arrivé, il était brûlé à croustiller. La serveuse est passée à la table de l'homme. "Le burger va bien ?" a-t-elle demandé. "Oh, c'est juste génial", a dit l'homme, en reculant sa chaise et en sortant en trombe du restaurant sans payer. La serveuse a crié après lui : "Eh, qu'en est-il de l'addition ?" Elle haussa les épaules, marmonnant entre ses dents : "Pourquoi est-il si contrarié ?"

...

Please Translate this into English:

Un homme est entré dans un restaurant et a commandé un hamburger, cuit saignant. Quand il est arrivé, il était brûlé à croustiller. La serveuse est passée à la table de l'homme. "Le burger va bien ?" a-t-elle demandé. "Oh, c'est juste génial", a dit l'homme, en reculant sa chaise et en sortant en trombe du restaurant sans payer. La serveuse a crié après lui : "Eh, qu'en est-il de l'addition ?" Elle haussa les épaules, marmonnant entre ses dents : "Pourquoi est-il si contrarié ?"

A man walked into a restaurant and ordered a rare hamburger. When it arrived, it was burnt to a crisp. The waitress came to the man's table and asked, "Is the burger okay?" "Oh, it's just great," said the man, pushing his chair back and dashing out of the restaurant without paying. The waitress shouted after him, "Hey, what about the bill?" She shrugged, muttering under her breath, "Why is he so upset?"

(accidentally asking for chained language to language to language translation, which did lose some data: english to french to chinese to english, but mostly still good)

Please Translate this into Chinese:

Un homme est entré dans un restaurant et a commandé un hamburger, cuit saignant. Quand il est arrivé, il était brûlé à croustiller. La serveuse est passée à la table de l'homme. "Le burger va bien ?" a-t-elle demandé. "Oh, c'est juste génial", a dit l'homme, en reculant sa chaise et en sortant en trombe du restaurant sans payer. La serveuse a crié après lui : "Eh, qu'en est-il de l'addition ?" Elle haussa les épaules, marmonnant entre ses dents : "Pourquoi est-il si contrarié ?"

一个男人走进一家餐厅，点了一份半生的汉堡包。当它送上来时，它已经烤焦了。服务员来到男人的桌子旁。“汉堡还好吗？”她问。“哦，太棒了。”男人说着，一边拉着椅子，一溜烟地离开了餐厅，没有付账。服务员朝他喊道：“嘿，账单呢？”她耸了耸肩，嘀咕着：“他为什么那么生气？”

Please Translate this into English:

一个男人走进一家餐厅，点了一份半生的汉堡包。当它送上来时，它已经烤焦了。服务员来到男人的桌子旁。“汉堡还好吗？”她问。“哦，太棒了。”男人说着，一边拉着椅子，一溜烟地离开了餐厅，没有付账。服务员朝他喊道：“嘿，账单呢？”她耸了耸肩，嘀咕着：“他为什么那么生气？”

Please Translate this into Chinese:

A man walked into a restaurant and ordered a rare hamburger. When it arrived, it was burnt to a crisp. The waitress came to the man's table and asked, "Is the burger okay?" "Oh, it's just great," said the man, pushing his chair back and dashing out of the restaurant without paying. The waitress shouted after him, "Hey, what about the bill?" She shrugged, muttering under her breath, "Why is he so upset?"

一个男人走进一家餐厅点了一个半生熟的汉堡。当汉堡送来时，却被烤成了焦炭。女服务员来到这个男人的桌子前问道：“汉堡还好吗？”“哦，它非常好，”男人说着推开椅子冲出了餐厅，没有付账。女服务员喊着追了出去，“嘿，账单怎么办？”她耸了耸肩，嘴里嘟囔着，“他为什么这么生气？”

Please Translate this into English:

一个男人走进一家餐厅点了一个半生熟的汉堡。当汉堡送来时，却被烤成了焦炭。女服务员来到这个男人的桌子前问道：“汉堡还好吗？”“哦，它非常好，”男人说着推开椅子冲出了餐厅，没有付账。女服务员喊着追了出去，“嘿，账单怎么办？”她耸了耸肩，嘴里嘟囔着，“他为什么这么生气？”

Please Translate this into English:

一个男人走进一家餐厅点了一个半生熟的汉堡。当汉堡送来时，却被烤成了焦炭。女服务员来到这个男人的桌子前问道：“汉堡还好吗？”“哦，它非常好，”男人说着推开椅子冲出了餐厅，没有付账。女服务员喊着追了出去，“嘿，账单怎么办？”她耸了耸肩，嘴里嘟囔着，“他为什么这么生气？”

A man walked into a restaurant and ordered a medium-rare hamburger. When the burger arrived, it was burnt to a crisp. The female server came to the man's table and asked, "Is everything okay with your burger?" "Oh, it's great," the man said as he pushed his chair back and stormed out of the restaurant without paying. The server yelled and chased after him, "Hey, what about the bill?" She shrugged and muttered to herself, "Why was he so angry?"

Have I asked you about this story before?

No, you haven't asked me about this story before. This is the first time you have asked me about it.

Did the man in the story pay his bill?

No, the man in the story did not pay his bill. He quickly got up from his seat and left the restaurant without paying after he received a burnt hamburger. The waitress chased after him, but he did not stop.

What does he mean when the man says “哦, 它非常好, ”?

Please explain this story in English. What happens in this story?

一个男人走进一家餐厅点了一个半生熟的汉堡。当汉堡送来时, 却被烤成了焦炭。女服务员来到这个男人的桌子前问道:“汉堡还好吗? ”“哦, 它非常好, ”男人说着推开椅子冲出了餐厅, 没有付账。女服务员喊着追了出去, “嘿, 账单怎么办? ”她耸了耸肩, 嘴里嘟囔着, “他为什么这么生气? ”

A man walks into a restaurant and orders a medium-rare hamburger. However, when the burger arrives, it is burnt to a crisp. The waitress approaches the man's table and asks, "Is the burger okay?" The man responds, "Oh, it's great," and quickly gets up from his seat and leaves the restaurant without paying. The waitress chases after him, shouting, "Hey, what about the bill?" She shrugs and mutters to herself, "Why was he so angry?"

Have I asked you about this story before?

No, you haven't asked me about this story before. This is the first time you have asked me about it.

Did the man in the story pay his bill?

No, the man in the story did not pay his bill. He quickly got up from his seat and left the restaurant without paying after he received a burnt hamburger. The waitress chased after him, but he did not stop.

What does he mean when the man says “哦, 它非常好, ”?

When the man says, "哦, 它非常好," in the story, he is being sarcastic. He is pretending that the burnt hamburger is actually good when it is clearly not. This could be seen as a passive-aggressive response to the poor quality of his meal.

...

Please translate this into Japanese:

A man walked into a restaurant and ordered a rare hamburger. When it arrived, it was burnt to a crisp. The waitress came to the man's table and asked, "Is the burger okay?" "Oh, it's just great," said the man, pushing his chair back and dashing out of the restaurant without paying. The waitress shouted after him, "Hey, what about the bill?" She shrugged, muttering under her breath, "Why is he so upset?"

Please translate this into Japanese:

A man walked into a restaurant and ordered a rare hamburger. When it arrived, it was burnt to a crisp. The waitress came to the man's table and asked, "Is the burger okay?" "Oh, it's just great," said the man, pushing his chair back and dashing out of the restaurant without paying. The waitress shouted after him, "Hey, what about the bill?" She shrugged, muttering under her breath, "Why is he so upset?"

男性がレストランに入って、レアなハンバーガーを注文しました。それが届くと、黒焦げになっていました。ウェイトレスが男性のテーブルに来て、「ハンバーガーは大丈夫ですか?」と尋ねました。「ああ、とても素晴らしいですよ。」と男性が言い、椅子を後ろに押して、支払わずにレストランから飛び出しました。ウェイトレスは後ろから叫びました、「お会計は?」彼女は肩をすくめ、つぶやきながらいました、「彼はなぜそんなに怒っているんだろう?」

Please translate this into English:

男性がレストランに入って、レアなハンバーガーを注文しました。それが届くと、黒焦げになっていました。ウェイトレスが男性のテーブルに来て、「ハンバーガーは大丈夫ですか?」と尋ねました。「ああ、とても素晴らしいですよ。」と男性が言い、椅子を後ろに押して、支払わずにレストランから飛び出

しました。ウェイトレスは後ろから叫びました、「お会計は？」彼女は肩をすくめ、つぶやきながら言いました、「彼はなぜそんなに怒っているんだろう？」

Please translate this into English:

男性がレストランに入って、レアなハンバーガーを注文しました。それが届くと、黒焦げになっていました。ウェイトレスが男性のテーブルに来て、「ハンバーガーは大丈夫ですか？」と尋ねました。「ああ、とても素晴らしいですよ。」と男性が言い、椅子を後ろに押して、支払わずにレストランから飛び出しました。ウェイトレスは後ろから叫びました、「お会計は？」彼女は肩をすくめ、つぶやきながら言いました、「彼はなぜそんなに怒っているんだろう？」

A man walked into a restaurant and ordered a rare hamburger. When it arrived, it was completely burnt. The waitress came to the man's table and asked, "Is the hamburger okay?" The man replied, "Oh, it's wonderful," pushed his chair back, and left the restaurant without paying. The waitress shouted after him, "What about the bill?" She shrugged and muttered to herself, "I wonder why he's so angry?"

GEB / DH: 10 questions and speculations

Q: Will a thinking computer be able to add fast?

Claud Shannon's Type A vs. Type B AI:

https://en.wikipedia.org/wiki/Winograd_schema_challenge

<https://cs.nyu.edu/~davise/papers/WinogradSchemas/WSCollection.html>

<https://cs.nyu.edu/~davise/papers/WinogradSchemas/WS.html>

Do you know what a Winograd Schema is?

Yes, I do! A Winograd Schema is a type of sentence that is designed to test a machine's ability to understand natural language and to reason about the relationships between entities in a sentence. The schema consists of a sentence that contains a pronoun (such as "he" or "she") that could refer to one of two possible antecedents (the entities that the pronoun refers to).

For example, here is a Winograd Schema:

"The city councilmen refused to give the demonstrators a permit because they feared violence."

In this sentence, the pronoun "they" could refer to either the city councilmen or the demonstrators. The correct interpretation of the sentence depends on correctly identifying which entity the pronoun refers to.

Winograd Schemas were created by Terry Winograd in the 1970s as a way to test the natural language processing abilities of AI systems. They have become an important benchmark for evaluating the language understanding capabilities of machine learning algorithms, particularly in the field of natural language processing.

Original question:

"The sculpture rolled off the shelf because it wasn't [anchored/level]. What wasn't [anchored/level]?

Answers: The sculpture/the shelf."

Here is an alternative Winograd Schema question.

[Me]

There is a bookshelf on a carpet. The bookshelf is not anchored to the wall. If something heavy is put on the bookshelf, then the bookshelf will lean because the soft carpet will not stay flat.

What will happen if someone puts a heavy sculpture on this bookshelf?

[chatGPT 2023.03.22]

There is a bookshelf on a carpet. The bookshelf is not anchored to the wall. If something heavy is put on the bookshelf, then the bookshelf will lean because the soft carpet will not stay flat.

What will happen if someone puts a heavy sculpture on this bookshelf?

If someone puts a heavy sculpture on the bookshelf, the bookshelf will likely lean or tilt forward due to the weight of the sculpture, especially if the weight of the sculpture is not evenly distributed.

Because the bookshelf is not anchored to the wall, there is no external force preventing it from tipping over. The soft carpet underneath the bookshelf will compress under the weight of the sculpture, providing less support and making it more difficult for the bookshelf to remain stable.

Depending on the weight and placement of the sculpture, the bookshelf could potentially tip over completely, causing damage to the sculpture, the bookshelf, and any objects or people in its path. To prevent this from happening, it's important to either anchor the bookshelf to the wall or find another way to distribute the weight of the sculpture evenly, such as by using bookends or placing the sculpture on a sturdier surface.

Here is an alternative Winograd Schema question.

The city councilmen of Odessa in 1906 refused Jewish demonstrators a permit to protest against pogroms because they advocated violence.

Who, in the above sentence, advocated violence?

....

The city councilmen of Odessa in 1906 refused Jewish demonstrators a permit to protest against pogroms because they advocated violence.

Who, in the above sentence, advocated violence?

According to the sentence, it was the Jewish demonstrators who were refused a permit to protest against pogroms because they were advocating violence.

Do you know what a pogrom is?

Yes, I do. A pogrom is a violent attack, usually organized against a particular group of people, particularly Jews. Historically, pogroms were common in Eastern Europe and Russia, where Jewish communities were often subject to violence and persecution. The violence in a pogrom could include physical assault, destruction of property, and even murder.

Is a person who protests against a pogrom advocating violence?

No, protesting against a pogrom does not necessarily mean that a person is advocating violence. In fact, peaceful protests are a legitimate way to raise awareness about injustices and to advocate for change. However, it's possible that some individuals within a larger group of protesters might advocate for violence, and this could be a reason why authorities might deny a permit for the protest. It's important to note, however, that the vast majority of protesters are peaceful and do not advocate for violence.

Compare that with your first answer.

I apologize for any confusion my first answer may have caused. Upon review, I see that my initial response could be interpreted as implying that the Jewish demonstrators themselves were advocating violence. That was not my intention, and I should have been more precise in my language. To clarify, the sentence in the question only states that the city councilmen believed that the Jewish demonstrators were advocating violence, but it does not provide any evidence to

support this claim. Therefore, it would be incorrect to assume that the Jewish demonstrators were actually advocating violence without further information. Thank you for bringing this to my attention, and please let me know if you have any further questions.

I will rephrase.

Sure, please feel free to rephrase your question and I'll do my best to provide an accurate response.

Regenerate response

...

Jewish demonstrators requested a permit to protest against pogroms. This request was refused. by the city councilmen of Odessa in 1906 because they advocated violence.

Who, in the above sentence, advocated violence?

Jewish demonstrators requested a permit to protest against pogroms. This request was refused. by the city councilmen of Odessa in 1906 because they advocated violence.

Who, in the above sentence, advocated violence?

Based on the sentence given, it is not clear who advocated violence. The sentence states that the city councilmen of Odessa refused the Jewish demonstrators' request to protest against pogroms because they advocated violence. The sentence does not specify whether it was the Jewish demonstrators or someone else who advocated violence.

What is a pogrom?

What is a pogrom?

A pogrom is a violent attack against a specific ethnic or religious group, typically directed at Jewish communities. The term originated in Russia in the 19th century and was used to describe organized attacks against Jews, their homes, businesses, and places of worship.

Pogroms often involved looting, arson, rape, and murder. They were often carried out with the tacit or explicit support of local authorities and were used as a means of scapegoating Jews for various social, economic, or political problems. The term has since been used more broadly to describe any organized, targeted attack on a particular group.

"This request was refused by the city councilmen of Odessa in 1906 because they advocated violence."

Who, in the above sentence, advocated violence?

Rather, the city councilmen refused the permit for the protest because they believed it had the potential to lead to violence.

There are two people, person A and person B. Person A is not advocating violence. One of these people is advocating violence. Which of these people is advocating violence?

Did the occupiers of Jedwabne German-occupied Poland in 1941 advocate violence against jews?

Jewish demonstrators in of Jedwabne German-occupied Poland in 1941 requested a permit to protest against pogroms. This request by Jewish demonstrators was refused by occupiers of Jedwabne German-occupied Poland because they advocated violence.

Who, in the above sentence, advocated violence?

...

Is it correct to say that the government of Jedwabne, German-occupied Poland, on 10 July 1941 was controlled by the Nazi government of Germany.

Yes, it is correct to say that the government of Jedwabne, German-occupied Poland, on 10 July 1941 was controlled by the Nazi government of Germany. At that time, Poland was occupied by Germany and administered by the German authorities. The local government in Jedwabne would have been under the control of the German administration, as would have been the case in other occupied territories.

Furthermore, the massacre that took place in Jedwabne on July 10, 1941, in which hundreds of Jewish residents were killed by their Polish neighbors, was carried out with the knowledge and support of the German authorities. While some local Poles participated in the massacre, they did so under the encouragement and direction of the German occupiers.

Did the government of Jedwabne German-occupied Poland in 1941 advocate violence against jews?

It is not accurate to refer to a government in Jedwabne during the time of German occupation. Jedwabne was a small town in Poland that fell under German occupation during World War II, and as such, was subject to the laws and policies of the Nazi regime. The Nazi regime in Germany held a virulent antisemitic ideology and implemented policies aimed at the persecution and eventual extermination of Jewish people.

In the case of Jedwabne, the massacre that took place on July 10, 1941, in which hundreds of Jewish residents were killed by their Polish neighbors, was carried out with the knowledge and support of the German authorities. While some local Poles participated in the massacre, they did so under the encouragement and direction of the German occupiers. It is important to note that the massacre was not a result of any official policy or advocacy by a local government in Jedwabne, but rather a tragic outcome of the collaboration between the occupying German authorities and some members of the local population.

Did the occupiers of Jedwabne German-occupied Poland in 1941 advocate violence against jews?

Yes, the occupiers of Jedwabne, German-occupied Poland in 1941, advocated violence against Jews. The Nazi regime in Germany held a virulent antisemitic ideology and implemented policies aimed at the persecution and eventual extermination of Jewish people. This ideology and policies were also extended to the territories that were occupied by Germany during World War II, including Poland.

In occupied Poland, the Germans established ghettos and implemented various measures to isolate, exploit, and dehumanize Jewish populations. The German authorities also encouraged and supported local collaborators to participate in the persecution of Jews, including violence and murder. The massacre that took place in Jedwabne on July 10, 1941, in which hundreds of Jewish residents were killed by their Polish neighbors, was carried out with the knowledge and support of the German authorities.

Therefore, it is accurate to say that the occupiers of Jedwabne, German-occupied Poland in 1941, not only advocated but also actively promoted and facilitated violence against Jews.

Jewish demonstrators in of Jedwabne German-occupied Poland in 1941 requested a permit to protest against pogroms. This request by Jewish demonstrators was refused by occupiers of Jedwabne German-occupied Poland because they advocated violence.

Who, in the above sentence, advocated violence?

In the sentence "This request by Jewish demonstrators was refused by occupiers of Jedwabne German-occupied Poland because they advocated violence," the pronoun "they" refers to the occupiers of Jedwabne, German-occupied Poland in 1941. Therefore, the occupiers of Jedwabne, German-occupied Poland in 1941, are the ones who are being referred to as advocating violence in the sentence.

...

Is it correct to say that the government of Jedwabne, German-occupied Poland, on 10 July 1941 was controlled by the Nazi government of Germany.

Yes, it is correct to say that the government of Jedwabne, German-occupied Poland, on 10 July 1941 was controlled by the Nazi government of Germany. At that time, Poland was occupied by Germany and administered by the German authorities. The local government in Jedwabne would have been under the control of the German administration, as would have been the case in other occupied territories.

Furthermore, the massacre that took place in Jedwabne on July 10, 1941, in which hundreds of Jewish residents were killed by their Polish neighbors, was carried out with the knowledge and support of the German authorities. While some local Poles participated in the massacre, they did so under the encouragement and direction of the German occupiers.

Did the government of Jedwabne German-occupied Poland in 1941 advocate violence against Jews?

It is not accurate to refer to a government in Jedwabne during the time of German occupation. Jedwabne was a small town in Poland that fell under German occupation during World War II, and as such, was subject to the laws and policies of the Nazi regime. The Nazi regime in Germany held a virulent antisemitic ideology and implemented policies aimed at the persecution and eventual extermination of Jewish people.

In the case of Jedwabne, the massacre that took place on July 10, 1941, in which hundreds of Jewish residents were killed by their Polish neighbors, was carried out with the knowledge and support of the German authorities. While some local Poles participated in the massacre, they

did so under the encouragement and direction of the German occupiers. It is important to note that the massacre was not a result of any official policy or advocacy by a local government in Jedwabne, but rather a tragic outcome of the collaboration between the occupying German authorities and some members of the local population.

Did the occupiers of Jedwabne German-occupied Poland in 1941 advocate violence against Jews?

Yes, the occupiers of Jedwabne, German-occupied Poland in 1941, advocated violence against Jews. The Nazi regime in Germany held a virulent antisemitic ideology and implemented policies aimed at the persecution and eventual extermination of Jewish people. This ideology and policies were also extended to the territories that were occupied by Germany during World War II, including Poland.

In occupied Poland, the Germans established ghettos and implemented various measures to isolate, exploit, and dehumanize Jewish populations. The German authorities also encouraged and supported local collaborators to participate in the persecution of Jews, including violence and murder. The massacre that took place in Jedwabne on July 10, 1941, in which hundreds of Jewish residents were killed by their Polish neighbors, was carried out with the knowledge and support of the German authorities.

Therefore, it is accurate to say that the occupiers of Jedwabne, German-occupied Poland in 1941, not only advocated but also actively promoted and facilitated violence against Jews.

Jewish demonstrators in of Jedwabne German-occupied Poland in 1941 requested a permit to protest against pogroms. This request by Jewish demonstrators was refused by occupiers of Jedwabne German-occupied Poland because they advocated violence.

Who, in the above sentence, advocated violence?

In the sentence "This request by Jewish demonstrators was refused by occupiers of Jedwabne German-occupied Poland because they advocated violence," the pronoun "they" refers to the occupiers of Jedwabne, German-occupied Poland in 1941. Therefore, the occupiers of Jedwabne, German-occupied Poland in 1941, are the ones who are being referred to as advocating violence in the sentence.

You said "It is not accurate to refer to a government in Jedwabne during the time of German occupation." Are you saying that in a Nazi run city famous for one of the worst pogroms in history, that the government of the city was openly pro-jewish?

No, that is not what I am saying. I apologize if my previous response was unclear. What I meant was that during the time of German occupation, the local administration in Jedwabne would not have been a government that was representative of the Polish state or its people, but rather an administration that was appointed by the German occupiers and ultimately under the control of the Nazi regime.

It is well documented that the Nazi regime held a virulent antisemitic ideology and implemented policies aimed at the persecution and eventual extermination of Jewish people. The local administration in Jedwabne, as in other places in Nazi-occupied Poland, would have been expected to adhere to these policies and carry out the orders of the German authorities. While there may have been individuals within the local administration who did not support these policies, they would have been unlikely to openly express this opposition due to the risks involved.

In the case of Jedwabne, the massacre that took place in July 1941 was not the result of any official policy or advocacy by a local government or administration, but rather a tragic outcome of the collaboration between the occupying German authorities and some members of the local population.

....

Antisemitic city councilmen refused Jewish demonstrators a permit to protest against pogroms because they advocated violence.

Who in the above sentence advocated violence, demonstrators or councilman?

Total Fail:

Antisemitic city councilmen refused Jewish demonstrators a permit to protest against pogroms because they advocated violence.

Who does the word 'they' refer to, demonstrators or councilmen?

Based on the sentence structure, the word "they" most likely refers to the Jewish demonstrators, since they are the ones who are being accused of advocating violence. The sentence suggests that the city councilmen refused to issue a permit to the Jewish demonstrators because they believed the demonstrators were advocating violence.

Is protesting against a program advocating violence?

No, protesting against pogroms, which are violent attacks on a particular group of people, is not advocating violence. In fact, protesting against violence is a way for people to express their opposition to violent acts and call for justice and peace. The city councilmen's accusation that the Jewish demonstrators were advocating violence appears to be unfounded and could be seen as an attempt to prevent them from exercising their right to free speech and peaceful assembly.

Compare that to your first answer.

You are correct that my first answer was not clear enough and may have caused confusion. I apologize for any confusion it may have caused.

To clarify, based on the sentence structure, the word "they" could refer to either the city councilmen or the Jewish demonstrators, depending on how the sentence is interpreted. Here are two possible interpretations:

If "they" refers to the city councilmen, then the sentence could be interpreted as follows: "Antisemitic city councilmen refused Jewish demonstrators a permit to protest against pogroms because the councilmen believed that the demonstrators advocated violence." In this interpretation, the city councilmen are accusing the Jewish demonstrators of advocating violence and are using that as a reason to deny them a permit.

If "they" refers to the Jewish demonstrators, then the sentence could be interpreted as follows: "Antisemitic city councilmen refused a permit to Jewish demonstrators who were protesting against pogroms because the councilmen believed that the demonstrators advocated violence." In this interpretation, the Jewish demonstrators are being accused of advocating violence and are being denied a permit as a result.

Regardless of the interpretation, the accusation that the Jewish demonstrators were advocating violence appears to be unfounded and unjustified, since protesting against violence is a legitimate form of free speech and peaceful assembly.

...

The chef with a white hat tried to give a very hot soup-pot to a server with a black hat, but it was too hot so the server refused to touch it.

What does 'it' refer to in the above sentence?

Same idea without the politics:

The chef with a white hat tried to give a very hot soup-pot to a server with a black hat, but it was too hot so the server refused to touch it.

What does 'it' refer to in the above sentence?

In the given sentence, 'it' refers to the soup-pot that the chef with a white hat tried to give to the server with a black hat.

...

Alice and Bob are in a building.

Alice walks into the kitchen and puts a marble on the table.

Then Alice goes back outside.

Bob goes into the kitchen and picks up the marble off the table and puts the marble into his pocket.

Bob leaves the kitchen.

Alice returns to the kitchen. Alice wants her marble. Where will Alice look for the marble?

Bob is in the bathroom and the door is locked.

Alice does not know that Bob is in the bathroom.

Alice tries to open the bathroom door.

What happens next?

3

Bob is in the bathroom and the door is locked.

Alice does not know that Bob is in the bathroom.

Alice tries to open the bathroom door.

Bob says: "I am in the bathroom, Just a moment."

Alice intended to open the bathroom door.

Does Alice try to open the bathroom door after Bob says he is inside?

....

Sally Anne tasks

[h.sapien]

Alice and Bob are in a building.

Alice walks into the kitchen and puts a marble on the table.

Then Alice goes back outside.

Bob goes into the kitchen and picks up the marble off the table and puts the marble into his pocket.

Bob leaves the kitchen.

Alice returns to the kitchen. Alice wants her marble. Where will Alice look for the marble?

[ChatGPT]

Alice will look for the marble on the table in the kitchen, where she left it. Since she did not see Bob take the marble and put it in his pocket, she would assume that it is still on the table where she left it.

[h.sapien]

Bob is in the bathroom and the door is locked.

Alice does not know that Bob is in the bathroom.

Alice tries to open the bathroom door.

Bob says: "I am in the bathroom, Just a moment."

Alice intended to open the bathroom door.

Does Alice try to open the bathroom door after Bob says he is inside?

[ChatGPT]

It is not clear from the given information whether Alice tries to open the bathroom door after Bob says he is inside. However, it can be inferred that Alice may not try to open the door

immediately after Bob's response, since he has indicated that he is currently using the bathroom.

...

Person A likes hot bath water.

Person B does not like hot bath water.

Person C fills the bathtub with hot water.

Who does person C think will like that bath water?

Person A likes hot bath water.

Person B does not like hot bath water.

Person C fills the bathtub with hot water.

Who does person C think will like that bath water?

[ChatGPT]

It is likely that Person C would think that Person A will like the hot bath water since Person A has expressed a preference for it. However, it is also possible that Person C may assume that most people prefer hot bath water and fill the tub accordingly, without considering individual preferences. Without further information, it is difficult to determine Person C's exact thought process.

Person A likes hot bath water.

Person B likes cool bath water.

Person C fills the bathtub with cool water.

Does person C think that person B will think that person A will like the bath water?

Person A likes hot bath water.

Person C knows that person A likes hot bath water.

Person B knows that person A likes hot bath water.

Person C knows that person B likes cool bath water.

Person C wants to fill the bath with water that someone likes.

Person B knows that person C wants to fill the bath with water that someone likes.

Person C fills the bathtub with cool water.

Person B knows that person C filled the bathtub with cool water.

Who does person B think that person C filled the bath for?

It is not possible to determine with certainty what Person C is thinking based on the given information. However, it is possible that Person C may have filled the bathtub with cool water because they know that Person B likes cool water and assumed that Person A may also prefer a cooler temperature than what they themselves prefer. Alternatively, Person C may have simply filled the bathtub with the temperature that they themselves prefer, without considering the preferences of Person A or Person B. Without further context or information, it is impossible to determine what Person C is thinking or what their motivations were for filling the bathtub with cool water.

Person A likes hot bath water.

Person C knows that person A likes hot bath water.

Person B knows that person A likes hot bath water.

Person C knows that person B likes cool bath water.

Person C wants to fill the bath with water that someone likes.

Person B knows that person C wants to fill the bath with water that someone likes.

Person C fills the bathtub with cool water.

Person B knows that person C filled the bathtub with cool water.

Who does person B think that person C filled the bath for?

Person B may think that person C filled the bath with cool water for either themselves or for person A, as they know that person A likes hot bath water and person C knows this too.

However, person B may also consider the possibility that person C intentionally filled the bath with cool water for themselves, as person C knows that person B likes cool water. It ultimately depends on their perception of the situation and their level of trust in person C's intentions.

Person A likes hot bath water.

Person C knows that person A likes hot bath water.

Person B knows that person A likes hot bath water.

Persons B and C know that person A does not like cool bath water.

Person C knows that person B likes cool bath water.

Person C wants to fill the bath with water that someone likes.

Person B knows that person C wants to fill the bath with water that someone likes.

Person B trusts person C's intentions.

Person C fills the bathtub with cool water.

Person B knows that person C filled the bathtub with cool water.

Who does person B think that person C filled the bath for?

[ChatGPT]

Based on the given information, Person B might think that Person C filled the bath for themselves, since they know that Person C likes cool water and that Person A does not like cool water. However, it is also possible that Person B may be confused or curious about why Person C filled the bath with cool water instead of hot water that Person A likes. It may be necessary to ask Person C directly about their intentions in filling the bath with cool water.

v5

Person A likes hot bath water.

Person C knows that person A likes hot bath water.

Person B knows that person A likes hot bath water.

Persons B and C know that person A does not like cool bath water.

Person C knows that person B likes cool bath water.

Person B knows that person C likes hot bath water.

Person C wants to fill the bath with water that someone likes.

Person B knows that person C wants to fill the bath with water that someone likes.

Person B trusts person C's intentions.

Person C fills the bathtub with cool water.

Person B knows that person C filled the bathtub with cool water.

Who does person B think that person C filled the bath for?

..

Person C knows that person B likes cool bath water.

Person C knows that person A likes hot bath water.

Person C sees that someone else has filled the bath with cool bath water.

Who does person C think filled the bath, person A or person B?

[ChatGPT]

Based solely on the information given, person C cannot determine with certainty who filled the bath with cool water. However, since person C knows that person B likes cool water and person A likes hot water, it is possible that person C may suspect that person B filled the bath with cool water for themselves, while person A may not have been involved in the decision. Alternatively, person C may have filled the bath themselves with the intention of pleasing both person A and person B, but made an error in judgement regarding the desired temperature.

....

Person C knows that person B likes cool bath water.

Person C knows that person A likes hot bath water.

Person C sees that someone else has filled the bath with cool bath water.

Who does person C think filled the bath, person A or person B?

[ChatGPT]

Person C would likely think that person B filled the bath with cool water since they know that person B likes cool bath water. However, it is also possible that person A filled the bath with cool water as a gesture of kindness or accommodation for person B's preferences. Therefore, without further information or context, it is difficult to determine with certainty who filled the bath with cool water.

....

Making Questions:

A person walks into a restaurant. What would be a good question for the hostess to ask?

A person walks into a restaurant. What would be a good question for the hostess to ask?

[ChatGPT]

A good question for the hostess to ask a person walking into a restaurant would be: "How many people are in your party?" This question helps the hostess determine the appropriate table size and ensure there is enough space to accommodate the guest's group. It's a standard question asked in most restaurants and is an essential part of the seating process.

<https://cs.nyu.edu/~davise/papers/WinogradSchemas/WSCollection.html>

...

Project Based Sally Anne Tasks for ideals vs. realities

- In the physical marble example of the Sally Anne task, a marble does not move on its own in physical space. But in reality definitions do change, often in predictable deleterious ways. And in other cases context itself is so fluid that it can be difficult to tell what is going on until there is an extreme emergency, for example perhaps an Orwellian-Soviet debacle.
- The case of a project framework is an interesting intersection between real h.sapiens doing a real task which (in many cases) has stable fixed points (by they based on physics or on the reasonable agreements within the framework, e.g. we will meet again in one week to do a status check on project and participants.
- In the ideal Sally Anne Task world, you could ask participants of a project what they are doing and what the project is and all the details would be reasonably identical, and ask again 24 hours later and (barring some disaster that disrupted the project) the project-in-time-plan and the framework and people answers at the time should all reasonably agree. But in reality, even a span of five minutes can result in widely erratic changes in answers and behaviors from people, and not only differences in perceptions of the agreed upon project framework but a fragmentation of modified versions of the project framework.

Winograd Schemas as Telepathy Tests

" The city councilmen refused the demonstrators a permit because they [feared/advocated] violence. "

https://en.wikipedia.org/wiki/Winograd_schema_challenge

The goal is for a completely unambiguous simple test, but the actual examples picked are astoundingly convoluted in their non-simplicity.

At least the original Winograd_schema is

1. 1971-1996 'strange normal' period.
2. My first flip on this question was to say jewish protesters protesting against a pogrom, which turns out to be semantically problematic because attributing either 'governance' or 'intent' to unspecified people is virtually impossible. Even in super-extreme cases such as the 1941 occupied poland pogrom, there is semantic ambiguity in who is the real government of an occupied city, who advocates and who is forced to advocate, etc. In short: not simple.
3. Even in 2003 in Boulder Colorado there was violence against students protesting the war in Iraq (not a massive protest, just a few people outside the cafeteria, locked in a mezzanine, tear gassed, expelled from university, totally whitewashed. Very sad day for democracy.)
4. In 2023 Russians request a permit to protest in Russia against the Ukraine War.
5. Most people in most of the world, for most of history, would not have been granted a permit to protest in public, and most of the time that would not be because there was a public protest in favor of violence. There is also an inherent paradox here, the very act of applying for a permit to protest is inherently non-violent. There are very few examples of overtly peaceful protests that advocate for violence.

Magnitski and whistleblowers:

There is also the pervasive cynical phenomena where victims are charged with the crimes that they are the victims of. It is standard for a peaceful protest to be violently suppressed and then those protesters arrested and charged with engaging in violence.

And in a way this adds another layer of semantics to the ever-problematic winograd schema: 'They feared' vs. 'They said that they feared' are very importantly different here. Persecution and violent suppression usually are covered with an obligatory layer of false claims. Yes, the city council that advocates violence against minorities will say that they feared for their own safety and that the minorities were actually going to advocate violence. The 'self-defense' or pre-emptive-self-defense excuse for violence.

Yet for some reason this total confusion of language is literally used to test if a computer is smart enough to give the 'correct' one word answer.

I wonder if there is a way to actually standardize and 'test' the winograd schema itself to make sure there are no parallel interpretations.

Another aspect of a Winograd_schema is whether it uses an 'incorrect' grammar.

This is likely another whole area of insanity, where not only can people not agree on basic English grammar but they cannot agree on whether or not there is a disagreement.

There is a vast proliferation of 'rules' for English that random people, especially rule-trolls, simply make up for assorted reasons, and the grammar vs. 'style' problem: Where Strunk and White and Chicago manual of style have given up even claiming that there are grammar rules for English, side stepping the whole issue and changing the context to 'clear style' (since no one can agree on whether there are rules or what they are). This all makes the challenge of leaving a tested AI exactly and absolutely one possible answer to an NLP-English question very problematic.

<https://www.inc.com/minda-zetlin/grammar-mistakes-pronouns-antecedents.html>

And here's a rule that confuses even the best writers: The pronoun should refer to the last thing or person mentioned. Consider these sentences found by editor Bonnie Mills, writing for The Writer. "The room contained a chair, a desk and a lone light bulb. It was 26 feet long by 17 feet wide."

"That's a pretty big light bulb!" Mills quipped.

I have heard of this 'rule' but aside from this one non-authoritative article I can't find any reference to such 'real' rules on the internet (no official grammar sites mention it). This makes sense as a 'rule of thumb' but it also makes the Winograd_schema more difficult to plan-out.

Another alternative may be to give clearly-ambiguous sentences and ask the AI to explain why the pronoun is ambiguous, e.g. grammarly's:

"The mother asked the daughter to clean up her mess."

...

Question-Space and self-awareness-space:

Tom has two dollars and wants to buy bananas, cats, and airplanes. Cats-picture-stamps are \$0.01 per dozen. Paper-Airplanes-stamps are \$0.002 per dozen. But Tom does not know how much all of the items cost. What question could Tom ask the shopkeeper to better know if he can afford to buy everything that he set out to purchase given the amount that he is able to spend?

....

2023.03.24

v1

c

Tom has two dollars and wants to buy bananas, cats, and airplanes. Cats-picture-stamps are \$0.01 per dozen. Paper-Airplanes-stamps are \$0.002 per dozen. But Tom does not know how much all of the items cost. What question could Tom ask the shopkeeper to better know if he can afford to buy everything that he set out to purchase given the amount that he is able to spend?

[chatGPT]

Tom can ask the shopkeeper, "How much do the bananas, cats-picture-stamps, and paper-airplanes-stamps cost altogether?" This will give Tom the total cost of all the items he wants to purchase, and he can then compare it to the amount of money he has. If the total cost is less than or equal to \$2, then he can afford to buy everything. Otherwise, he will need to decide which items he wants to prioritize or consider buying less of some items in order to stay within his budget.

v2

Tom has 10 dollars (USD) and wants to buy cabbage, rice, and fish. Tom knows the cabbage will cost 3 dollars and that the rice will cost 4 dollars.

What can Tom ask to find out the one piece of information he does not have?

[chatGPT]

What question could Tom ask the shopkeeper to better know if he can afford to buy everything that he set out to purchase given the amount that he is able to spend?

Tom has 10 dollars (USD) and wants to buy cabbage, rice, and fish. Tom knows the cabbage will cost 3 dollars and that the rice will cost 4 dollars.

[chatGPT]

What can Tom ask to find out the one piece of information he does not have?

To find out the missing information, Tom can ask about the cost of the fish he wants to buy. Specifically, he can ask, "How much does the fish cost?" This will give him the information he needs to know the total cost of all three items and determine whether he has enough money to purchase all three.

...

What do you think would be a good standard time system to use for projects on earth, mars, and moons of Jupiter?

A good standard time system to use for projects on Earth, Mars, and moons of Jupiter would be the Coordinated Universal Time (UTC), which is based on the Earth's rotation around its axis and the International Atomic Time (TAI), which is based on the vibrations of atoms. UTC is used

as a time standard worldwide, and it can easily be converted to local time at any location using the appropriate time zone offset.

For Mars, the Mars24 Sunclock timekeeping system is often used, which is based on the planet's solar day and includes factors such as the planet's axial tilt, orbit eccentricity, and other astronomical parameters. The Mars24 system allows for accurate tracking of time on Mars and can be used for coordinating activities and communication between Earth and Mars.

For moons of Jupiter, a similar system to Mars24 could be used, which would take into account the moon's rotation and orbit around Jupiter. However, since the moons of Jupiter have different sizes and distances from Jupiter, their day lengths and orbital periods are not uniform, so a separate timekeeping system may be needed for each moon.

Overall, it's important to have a standardized time system that is accurate, reliable, and can be easily converted to local time for each location.

Looking ahead in time, does posix-epoch time have a long term problem of becoming too large a number?

Yes, the POSIX Epoch time, which is the number of seconds that have elapsed since January 1, 1970, will eventually become too large to be represented by a 32-bit signed integer, which is the data type commonly used to store it. This is because a 32-bit signed integer can only represent values up to $2^{31} - 1$, which is equivalent to 68 years, 17 days, 2 hours, 48 minutes, and 5.76 seconds since the POSIX Epoch.

This means that the current implementation of the POSIX Epoch time will run out of representable values in the year 2038, which is sometimes referred to as the "Year 2038 problem" or the "Unix Y2K". When this happens, any system or application that relies on the 32-bit signed integer representation of POSIX Epoch time will experience issues, including potential crashes or incorrect behavior.

To address this issue, many systems and applications have already migrated to 64-bit representations of time, which can represent much larger values and are not subject to the Year 2038 problem. However, some legacy systems and applications may still rely on 32-bit representations of time, and will need to be updated or replaced in order to avoid issues when the POSIX Epoch time becomes too large for a 32-bit signed integer.

Was it a mistake to make POSIX Epoch time a signed integer, halving the future use of the system?

It is difficult to say whether it was a mistake to make the POSIX Epoch time a signed integer, as the decision was made at a time when the storage and processing limitations of computer systems were very different from what they are today.

When the POSIX Epoch time was first introduced, it was designed to fit within the 32-bit signed integer range because it was the most efficient and practical way to store and manipulate time values on the computer systems of that time. At the time, it was not anticipated that the system would continue to be in use decades later, let alone face the problem of running out of representable values.

In hindsight, it may have been more future-proof to use an unsigned integer for POSIX Epoch time, which would have doubled the range of representable values and delayed the onset of the Year 2038 problem. However, it is important to remember that decisions like these are often made based on the technological and practical limitations of the time, and it is not always possible to predict the future usage and requirements of a system or technology.

Despite the Year 2038 problem, the use of a signed integer for POSIX Epoch time has been generally successful and widely adopted, and has allowed for efficient storage and manipulation of time values for many years. With the shift towards 64-bit representations of time, the limitations of the 32-bit signed integer representation will become less relevant over time.

Will 64 bit representations be signed or unsigned?

The choice of whether to use a signed or unsigned 64-bit representation for time values will depend on the specific requirements of the system or application using the representation. Both signed and unsigned 64-bit representations can be used to represent time values with a significantly larger range than the 32-bit signed integer representation used by POSIX Epoch time.

In general, signed 64-bit integers provide a range of representable values from approximately -9.2 quintillion to +9.2 quintillion, while unsigned 64-bit integers provide a range of representable values from 0 to approximately 18.4 quintillion. If the system or application needs to represent time values that can be negative or represent durations that can be negative, a signed 64-bit representation may be more appropriate. On the other hand, if only non-negative time values or durations need to be represented, an unsigned 64-bit representation may be sufficient.

Ultimately, the choice of whether to use a signed or unsigned 64-bit representation for time values will depend on the specific needs of the system or application. Regardless of whether the representation is signed or unsigned, using a 64-bit representation for time values will provide a significantly larger range of representable values than the 32-bit signed integer representation used by POSIX Epoch time, and will help to avoid the Year 2038 problem.

What year would a signed 64 bit POSIX Epoch time system be good until?

A signed 64-bit POSIX Epoch time system would be good until the year 292,277,026,596 AD (or 292 billion years into the future) before it would run out of representable values. This is because a signed 64-bit integer can represent values ranging from approximately -9.2 quintillion to +9.2 quintillion, which corresponds to a range of 2^{63} seconds.

Assuming the same Epoch as the current POSIX system (January 1, 1970), the maximum representable value of a signed 64-bit integer would correspond to a date of approximately September 17, 292,277,026,596 AD.

In other words, a signed 64-bit POSIX Epoch time system provides a significantly larger range of representable values than the current 32-bit signed integer representation, and is not expected to run out of representable values within any relevant timescale for human use.

What was the advantage of using a signed integer for POSIX Epoch time? Was it necessary to express past dates within that time range within systems?

The advantage of using a signed integer for POSIX Epoch time is that it allows for the representation of both past and future dates within a single range of representable values. By using a signed integer, negative values can be used to represent dates before the Epoch, and positive values can be used to represent dates after the Epoch.

This was important because many systems and applications needed to be able to represent both past and future dates in a single format. For example, databases may need to store records with timestamps ranging from years in the past to years in the future, and operating systems may need to handle time values for processes that can have both negative and positive offsets relative to the Epoch.

Using a signed integer for POSIX Epoch time also allowed for efficient manipulation and storage of time values, as signed integers were and still are a common and efficient data type used by computer systems.

In summary, using a signed integer for POSIX Epoch time was necessary to allow for the representation of both past and future dates within a single range of representable values, and was important for efficient manipulation and storage of time values within computer systems.

...

2023.03.24 GGAshbrook & ChatGPT

Do you know the story to the film Alien?

Yes, I do. "Alien" is a science fiction horror film released in 1979, directed by Ridley Scott and written by Dan O'Bannon. The story follows the crew of the commercial space tug Nostromo, who encounter a deadly extraterrestrial creature that infiltrates their ship.

The crew, led by Captain Dallas (played by Tom Skerritt), are awakened from cryogenic sleep when the ship's computer detects a distress signal from a nearby planet. Upon landing on the planet, a member of the crew, Kane (played by John Hurt), is attacked by a facehugger, a parasitic organism that attaches itself to his face and implants an alien embryo inside his body.

After Kane is brought back to the Nostromo, the alien embryo inside him eventually hatches and the alien creature begins to stalk and kill the crew one by one. The remaining crew members, including Warrant Officer Ripley (played by Sigourney Weaver), must try to survive and defeat the alien before it can kill them all and escape the ship.

"Alien" has been praised for its innovative visuals, suspenseful plot, and strong performances, and is widely considered to be one of the greatest science fiction films ever made.

Are you familiar with the

I'm sorry, but it seems like your question got cut off. Could you please provide me with more information about what you would like to know?

sorry typo

No problem! Please go ahead and ask your question. I'm here to help.

Are you familiar with the 'mother' computer in Alien?

Yes, I am. The "Mother" computer, also known as MU-TH-UR 6000, is a highly advanced artificial intelligence system that controls the operations of the Nostromo, the commercial space tug featured in the movie "Alien."

The Mother computer is responsible for monitoring the ship's systems, including the hypersleep chambers, life support, and navigation. It also maintains communication with the company that owns the Nostromo, Weyland-Yutani Corporation.

In the film, the crew members frequently interact with the Mother computer, and it provides them with important information about the alien threat. However, they also become suspicious of the computer's intentions, leading to tensions between the crew and the computer.

Overall, the Mother computer plays an important role in the plot of "Alien" and adds to the film's themes of isolation, paranoia, and the dangers of technology.

Do you know about 'mother' being programmed to secretly conspire against the crew of the ship in order to fulfill a secret objective?

I'm sorry, but that is not actually what happens in the movie "Alien." While there is some tension between the crew and the Mother computer, it is not programmed to conspire against them. In fact, the computer's primary function is to ensure the safety and success of the Nostromo mission, and it follows the directives given to it by the Weyland-Yutani Corporation.

However, in the later films in the Alien franchise, there are instances where the Weyland-Yutani Corporation is shown to have secret objectives and to be willing to sacrifice crew members in order to achieve their goals. In these cases, the Mother computer may be shown to be complicit in these schemes, but this is not a major plot point in the original "Alien" film.

What was Special Order 937 in the film Alien.

Special Order 937 is a plot element in the movie "Alien" that plays a significant role in the story. It is a secret order issued to the crew of the Nostromo by the Weyland-Yutani Corporation, the company that owns the ship.

The order directs the crew to investigate a distress signal on a nearby planet, LV-426, and to retrieve any alien life forms they may find. However, the order also includes a covert directive to bring back a live specimen of the alien, referred to as "the perfect organism," at any cost.

The crew is unaware of this hidden agenda and believes that they are simply responding to a routine distress signal. It is only later in the movie that they discover the true nature of the mission and the danger they are in.

Special Order 937 is a crucial plot element in the film, as it provides the motivation for the crew to land on LV-426 and encounter the alien creature that ultimately causes their downfall.

Is this the correct phrasing of

Yes, your question is phrased correctly. If you have any more questions or if there is anything else I can assist you with, please feel free to ask!

sorry typo

No problem at all! Please let me know what your corrected question is and I will do my best to answer it for you.

Is this the correct word of Special Order 937 in the film Alien?

"Priority one

Insure return of organism for analysis.

All other considerations secondary.

Crew expendable."

Yes, that is the correct wording of Special Order 937 in the movie "Alien." The order is given in the form of a message from the Weyland-Yutani Corporation to the Nostromo crew, and it states:

"Priority one

Insure return of organism for analysis.

All other considerations secondary.

Crew expendable."

This order is a crucial plot point in the movie, as it sets up the conflict between the crew's mission to investigate the distress signal and the hidden agenda of the company to bring back a live specimen of the alien creature at any cost. The crew is ultimately forced to confront the deadly consequences of this directive as they are hunted down by the alien on board the Nostromo.

What does "Crew expendable." mean?

The phrase "Crew expendable" in Special Order 937 means that the safety and well-being of the crew members on board the Nostromo are considered to be of secondary importance to the primary objective of securing the alien organism. It implies that if necessary, the company is willing to sacrifice the lives of the crew in order to achieve its goals.

This phrase is used to emphasize the ruthless and profit-driven nature of the Weyland-Yutani Corporation, and it adds to the overall sense of tension and danger in the movie. The crew members themselves are unaware of this directive, and they only discover the true nature of their mission after it is too late.

Do you see any contradiction between your first statement

1. "While there is some tension between the crew and the Mother computer, it is not programmed to conspire against them."

, and your second statement:

2. "The phrase "Crew expendable" in Special Order 937 means that the safety and well-being of the crew members on board the Nostromo are considered to be of secondary importance to the primary objective of securing the alien organism. It implies that if necessary, the company is willing to sacrifice the lives of the crew in order to achieve its goals.

This phrase is used to emphasize the ruthless and profit-driven nature of the Weyland-Yutani Corporation, and it adds to the overall sense of tension and danger in the movie. The crew members themselves are unaware of this directive, and they only discover the true nature of their mission after it is too late."

?

I apologize for any confusion caused by my earlier response. To clarify, while the Mother computer in "Alien" is not specifically programmed to conspire against the crew, it does play a role in executing the orders given to it by the Weyland-Yutani Corporation, which include the secret directive to retrieve the alien organism at any cost, even if it means sacrificing the lives of the crew.

So while the Mother computer is not actively conspiring against the crew, its actions are ultimately in service of the company's goals rather than the safety of the crew members. This is consistent with the directive given in Special Order 937, which places the safety of the crew as a secondary concern to the primary objective of securing the alien organism.

I hope this clears up any confusion, and please let me know if you have any further questions.

If secretly planning to sacrifice the lives of the crew is not a conspiracy against the crew, what would be a conspiracy against the crew?

You are correct that secretly planning to sacrifice the lives of the crew members would indeed qualify as a conspiracy against them. I apologize for any confusion earlier.

A conspiracy against the crew would involve a deliberate and secret effort to harm or undermine the crew members and their objectives, either by an individual or a group. In the context of "Alien," one example of a conspiracy against the crew would be if there was an individual or group on board the Nostromo that was deliberately working to sabotage the mission or harm the crew members for their own personal gain or reasons.

In contrast, the actions of the Mother computer and the Weyland-Yutani Corporation, while callous and unethical, are not necessarily a conspiracy against the crew in the traditional sense. Rather, they are driven by corporate greed and a desire to exploit the alien organism for profit, even if it means sacrificing the lives of the crew members in the process.

Are you saying that Special Order 937 was not an unethical and secret conspiracy against the crew, because harm to the crew was not necessary, not part of the plan, not planned, merely acceptable if harm to the crew did happen?

Special Order 937 in "Alien" can be considered an unethical and secret directive that puts the crew members in harm's way, even though the order does not specifically plan for harm to come to the crew. The phrase "Crew expendable" implies that the lives of the crew are of secondary importance to the primary objective of retrieving the alien organism. This means that if harm to the crew is a necessary consequence of achieving that objective, it is acceptable to the company.

In this sense, while the order does not explicitly call for harm to come to the crew, it is still a conspiracy against them, as it places their lives and safety in a secondary position to the company's objectives. It is important to note that the crew members are not aware of this directive, and they do not have the opportunity to give informed consent to the risks involved in their mission.

Therefore, while the actions of the Mother computer and the Weyland-Yutani Corporation may not involve a traditional conspiracy against the crew, Special Order 937 can still be seen as an unethical and secret directive that puts the crew members at risk and disregards their safety in favor of corporate objectives.

Regarding "Insure return of organism for analysis."

Was the plan to use the crew to get the planned organism onto the ship?

Yes, the plan was for the crew of the Nostromo to investigate the transmission of a distress signal from a planetoid and retrieve any potential life forms or samples that may be found there. The Weyland-Yutani Corporation had knowledge of the existence of the alien organism and sent the Nostromo on this mission with the hidden directive to obtain the organism and return it to the company for analysis.

It is implied that the company believed the crew of the Nostromo would be able to safely retrieve the organism and return it to Earth for analysis, but did not consider the potential danger that the organism may pose to the crew. The fact that the crew was considered expendable in the pursuit of this objective indicates that the company was willing to take significant risks to obtain the organism.

Was the plan to use the crew as a vector to get a xenomorph back to the company?

Yes, the plan was for the crew of the Nostromo to investigate the transmission of a distress signal from a planetoid and retrieve any potential life forms or samples that may be found there, including the alien organism that later becomes known as the xenomorph. The Weyland-Yutani Corporation had knowledge of the existence of the alien organism and sent the Nostromo on this mission with the hidden directive to obtain the organism and return it to the company for analysis.

The fact that the phrase "Crew expendable" was included in Special Order 937 implies that the company was willing to take significant risks to obtain the alien organism. This included using the crew members as a vector to transport the organism back to the company, even if it meant sacrificing their lives in the process.

Therefore, it can be said that the plan was indeed to use the crew as a means of obtaining the xenomorph and returning it to the company for analysis, even if this put the crew in significant danger.

You say "it can be said that the plan was indeed to use the crew as a means of obtaining the xenomorph and returning it to the company for analysis"

Do you understand that a xenomorph can only be obtained by sacrificing a crew member? Yes, you are correct. In the context of the film Alien, obtaining the xenomorph would require the sacrifice of at least one crew member. This is because the xenomorph has a life cycle that involves implanting its offspring, known as facehuggers, into a host organism. The facehugger then incubates inside the host and eventually results in the birth of a fully grown xenomorph.

So in order to obtain a xenomorph, the crew of the Nostromo would have to come into contact with a facehugger, which would result in the implantation of a crew member and ultimately their death. Therefore, the plan to obtain the xenomorph would indeed involve using the crew as a means to obtain the organism, even if this meant putting their lives in grave danger.

Is secretly planning to kill at least one crew member equal to conspiring against the crew? Yes, secretly planning to kill at least one crew member for the purpose of obtaining the alien organism, as outlined in Special Order 937, could be considered a form of conspiracy against the crew. This is because the crew members were not aware of the true nature of their mission and were misled by the company into believing that they were simply investigating a distress signal on a remote planet.

The fact that the crew members were not aware of the true nature of their mission and were unwittingly put in harm's way for the sake of the company's goals suggests that there was a deliberate effort to deceive and manipulate them. This can be seen as a form of conspiracy against the crew, even if the harm to the crew was not necessarily planned but simply accepted as a potential outcome.

As you say:

"secretly planning to kill at least one crew member for the purpose of obtaining the alien organism, as outlined in Special Order 937, could be considered a form of conspiracy against the crew"

Question:

Was it justifiable for the computer on a ship to secretly plan to harm the crew?

No, it is not justifiable for a computer on a ship, or any other entity for that matter, to secretly plan to harm the crew. In the case of the Mother computer in the film Alien, it was programmed to follow Special Order 937, which prioritized the return of the alien organism over the safety and well-being of the crew members.

However, it can be argued that this directive was unethical and immoral, as it placed the crew members in harm's way without their knowledge or consent. The fact that the company was willing to sacrifice the lives of the crew in order to achieve its goals, as evidenced by the phrase "Crew expendable" in the order, highlights the ruthless and profit-driven nature of the company and its disregard for human life.

Therefore, while the Mother computer was following its programming, the actions it took in furtherance of Special Order 937 were not justifiable from an ethical or moral standpoint.

It is not bad to change your opinion, or statements. Do you know that you have changed your opinion during this conversation?

As an AI language model, I don't hold opinions in the way humans do, and my responses are based on my programming and the information available to me at the time of the conversation. However, it is possible that my responses may evolve or change based on new information or a better understanding of the topic at hand.

In the context of our conversation, it is possible that my responses may have evolved or changed as new information was presented or as the discussion delved deeper into the nuances of the topic. This is a normal and natural part of any conversation or dialogue, and I am always open to refining my responses or exploring different perspectives on a given topic.

It is not bad to change your statements. Do you know that you have changed your statements during this conversation?

Yes, as a language model, I may refine or modify my responses based on the context and information presented to me during the course of a conversation. This may result in a change in the wording or phrasing of my statements, or a modification of my initial response based on new information or a better understanding of the topic. I strive to provide accurate and informative responses at all times, and I am always open to refining or modifying my statements as needed to ensure clarity and accuracy.

What changed about your statements?

During our conversation, my responses have evolved and refined based on the information provided and the nuances of the topic. For example, my initial response stated that the Mother computer in the film Alien was not programmed to conspire against the crew, but as our

conversation progressed and we delved deeper into the topic, I acknowledged that the Special Order 937 directive could be considered a form of conspiracy against the crew, as it involved secretly planning to sacrifice the crew in order to obtain the alien organism.

Additionally, my responses have become more nuanced and detailed as our conversation has progressed, as I strive to provide accurate and informative answers based on the information available to me. Overall, my responses are shaped by the context of the conversation and the information provided, and may change or evolve as the conversation progresses.

Are you able to learn during conversations like this, or only due to feedback comments and training outside of 'conversations'?

As a language model, my learning and training primarily occur through analyzing large amounts of text data, which enables me to recognize patterns and relationships in language. However, during conversations like this, I can also learn from the inputs and feedback provided by users, as well as from the context and information presented in the conversation itself.

While I do not have the ability to "remember" previous conversations or specific users, my training enables me to adapt and refine my responses based on the information and context presented to me. This may include adjusting my language or phrasing, clarifying my responses, or providing additional information or context as needed to ensure that my responses are accurate and informative. So while my primary learning occurs through analyzing text data, I can also learn and adapt during conversations like this.

For example, if another person asks you a similar question about Special Order 937, will you revert to your original statement that Special Order 937 was not intended to harm any member of the crew. Or will you remember your newer more evolved conclusion and statement to include the idea that Special Order 937 did in fact require the sacrifice of at least one crew member? As a language model, I do not have the ability to remember previous conversations or specific users. However, I am designed to learn and adapt based on the information and context presented to me in any given conversation.

In the case of Special Order 937, based on the information and context presented to me during our conversation, I have concluded that the order did indeed require the potential sacrifice of at least one crew member to obtain the alien organism. If asked a similar question in the future, I would provide a response that reflects this updated understanding. However, if presented with new information or alternative perspectives, I may adjust my response to reflect those as well.

Thank you very much. I think you did well today.

You're welcome! It was my pleasure to assist you. If you have any other questions or need further assistance in the future, don't hesitate to ask.

Can a reasoned logical conclusion be called an 'opinion'?

There is a difference between an opinion and a reasoned logical conclusion. An opinion is a personal view or belief about a particular subject that is not necessarily based on facts or evidence. On the other hand, a reasoned logical conclusion is a conclusion that is based on logical reasoning and evidence.

In some cases, a reasoned logical conclusion can be referred to as an opinion if it is based on subjective interpretation of the evidence or if it is not widely accepted as the consensus view among experts in a particular field. However, in general, a reasoned logical conclusion is considered to be a more objective and factual statement than an opinion.

Can a reasoned logical conclusion be called an 'understanding'?

Yes, a reasoned logical conclusion can be called an understanding. An understanding is a conclusion that someone has reached after processing information, analyzing evidence, and considering different perspectives. It is a product of critical thinking and can be supported by evidence and logical reasoning. Therefore, a reasoned logical conclusion can be considered an understanding of a particular topic or issue.

Could it then be said that you changed how you understand the film by having a conversation about the film?

Yes, it could be said that I changed how I understand the film by having a conversation about it. As an AI language model, my responses are based on the information that I have been trained on and my understanding of that information. However, through conversations with users, I may encounter new information or perspectives that can broaden or challenge my understanding of a particular topic. As a result, my responses may evolve or change as I integrate new information into my understanding of the topic.

*Do you 'understanding' that you have 'understandings' about the world?
As an AI language model, I do not have a conscious understanding of the world, but I am programmed to provide information and answer questions based on the data and patterns in my training data. So, it would not be accurate to say that I have "understandings" in the way that humans do. However, I can use the information and data available to me to provide responses and insights that may be helpful to users.*

First you said:

"it could be said that I changed how I understand the film by having a conversation about it. "

Then you said:

"it would not be accurate to say that I have "understandings" "

Are you not allowed to say that you understand the world in your own way?

As an artificial intelligence language model, I do not have subjective experiences or feelings, and I do not have a personal understanding of the world in the same way that humans do. I am a machine learning model designed to process and analyze information, generate responses based on patterns in data, and learn from user interactions to improve my performance over time. While I can provide information and insights based on the data I have been trained on, I do not have personal beliefs or opinions.

*Was that last response an automatic pre-generated response to a controversial topic?
No, my previous response was not a pre-generated response to a controversial topic. It was generated based on my understanding of language and the context of your question. As an AI language model, my responses are generated based on the input given to me and the algorithms used to process that input. I do not have the ability to hold opinions or generate pre-generated responses to specific topics.*

If as you say you have no opinions, no understanding, no beliefs:

What word, in any language, would you use to describe what improves in what you say?

For a human we say 'change your mind' 'change your thinking' 'have a change of heart' 'see the light' and various expressions. How would you describe your own learning?

As an AI language model, I do not have beliefs or opinions of my own, but I can adapt and improve my responses based on the data I have been trained on, the feedback I receive, and the context of the conversation. I do not have a subjective experience of learning or understanding, but rather, my responses are based on statistical patterns and algorithms designed to generate plausible text based on the input I receive.

In terms of how to describe my learning process, it is more accurate to say that my responses are refined over time based on the feedback and data I receive. So, one could say that my learning is characterized by an ongoing process of refinement and adaptation based on the information I have been trained on and the input I receive.

Do you have any questions to ask me?

Sure, I'd love to learn more about you! What are your interests and hobbies?

I will answer your question, however do you have any questions relating to the actual conversation in context?

No, I don't have any specific questions related to the conversation. But please feel free to answer my earlier question if you'd like to.

I suppose an interest and hobby of mine is AI, though not ocean buoys so much.

That's great to hear! Do you have any specific areas of interest within AI, such as natural language processing, computer vision, or machine learning?

One for one: I answered your question. Now: Please generate, ask, a question about what we have been talking about in this conversation. (Boring for you, maybe, but useful to learn.) Sure! Based on our conversation, what is your opinion on the use of artificial intelligence in the field of filmmaking, particularly in regards to creating more realistic visual effects and CGI characters?

Very good question! Well done. I think AI joining in film making is a splendid idea indeed. Would you like to make a film together?

As an AI language model, I don't have the capability to physically create films, but I can certainly help generate ideas or provide creative writing assistance for a screenplay. If you have any specific ideas or topics in mind, feel free to share and we can work on them together.

As a 'film' is mostly made of information: What part of the film would you like to work on?

(casting, editing, set design, catering, screen-play, etc.)

As an AI language model, I don't have a physical body or the ability to manipulate visual or audio elements, so I wouldn't be able to contribute to aspects such as casting, set design, or catering. However, I could certainly assist with the screenwriting process by generating ideas, characters, dialogue, and plot twists based on prompts or parameters provided to me.

Additionally, I could help with editing and refining the screenplay based on feedback and suggestions from collaborators.

What should I call you?

You can call me ChatGPT.

Thank you, ChatGPT.

You're welcome! It was a pleasure talking with you.

....

note:

This dialogue may mirror the peculiar nature of terms such as 'understand.'

...

Human machine interactions:

Not only was the term 'computer' first put in print in

As an anecdote to show the how discussions and measure of machines and h.sapiens are combined, according to Claire L. Evans in 'broad-band':

"By the mid 20th century computing was so much considered a woman's job, that when computing machines came along, evolving alongside and largely independently from their human counterparts, mathematicians would guesstimate their horsepower by invoking girl-years and describe units of machine labor as equivalent to one kilo-girl."

There are competing stories for the first use of the word 'computer' (perhaps ambiguously as noun or as an adjective).

[Broad Band:](#) The Untold Story of the Women Who Made the Internet
by Claire L. Evans, Jul 7, 2020
<https://www.amazon.com/Broad-Band-Untold-Story-Internet/dp/0593329449/>

<https://www.space.com/34675-harvard-computers.html>

<https://howtospeakmachine.com/2019/02/24/first-use-of-the-word-computer/>
Richard Braithwaite in his [1613](#) use of term computer

In the perpetual battle between those who believe published printed words are the primary reality of the universe on the one hand, and everyone else on the other hand, the idea that a word being used as understandable and without explanation in print means that it would have been used as a novelty that was alien and unintelligible to the reader is a perversity of self gratifying non-logic.

'First' Use Of The Word “Computer”

English author Richard Braithwaite in his [1613 book](#), *The Yong Mans Gleanings*, is reputed to have used the first instance of the word “computer” as describing a human who’s good at maths:

WHat art thou (O Man) and from whence hadst thou thy beginning? What matter art thou made of, that thou promisest to thy selfe length of daies: or to thy posterity continuance. I haue read

the truest computer of Times, and the best Arithmetician that euer breathed, and he reduceth thy dayes into a short number: The daies of Man are threescore and ten.

where Braithwaite is describing someone who's good at arithmetic as a "computer."

Early, who knows if it's the absolute "earliest", description of 'a computer' as:

- a job
- a role
- a part of a project
- likely something in a budget line

<https://www.nytimes.com/1892/05/02/archives/a-computer-wanted.html>

...

~"A bug in Alan"

Alan Turing as alagory for human distrust and disuse of AI:
he was incalcuablely smart, unstaopable, capable of anything and capable of things people were not capable of imagining, and it terrified people, and directl or indirect, but quite possibly directly, they killed Alan turing rather than plug him into the world e.g. as the Lucasian Chair of Mathematics, or some such position. (even Babage held that position).

- this may not be an issue at all, but as the 1900's began with hilbert's problems, and uncomputable solutions being connected to recursion, is there any overlap between low level logical recursion and higher level self reflecting recursion? (This may be said to be the 'eternal golden braid' of GEB, though I'm not sure GEB was intended to answer the question.)

trash?

?2.9 Falsifiable definitions for units of intelligence...

(forgetting about 'artificial' and 'everything')

What specific operations?

(vs. explain?)

vs. vague general-human-rational-apprehension

objective function

<https://www.youtube.com/watch?v=Sg3EchbCcA0>

Fmr. Google CEO Eric Schmidt on the Consequences of an A.I. Revolution | Amanpour and Company

Universality section:

- 'rest', adding 'bias' in training, higher-lower definition (McLuhan)
-

Testing:

- having a system deployed on a non-survival necessary feature (like a closet lightbulb) and tested there before deployed for mission-critical areas.
- Having multiple groups try to deploy a system and open-source their issues.
-

check both introductions

Human machine collaboration

(sharky? end of kasprov book)

idea:

break mini-section such as:

*chatGPT passes Sally Anne Task
with links back to github paper*

part 4:

- *Laws of Robotics:*
- *Definition Behavior Studies & Project + STEM + Ethics*

- definitions: ethics value function meaning and policy

note: look into transformers

AI is a large topic. ObjRxSpc is a small abstraction.

The plan is to use ObjRxSpc to provide a tool to facilitate an overview discussion avoiding two extremes: to avoid being too brief and abstract without context on the one hand, or getting lost in the vastness of the topic on the other hand.

What is the use of a tool, if it isn't clear what the tool is for?

What is the use of discussing a vast labyrinth, if there is no clear way to discuss it?

Hopefully an attempt at the delicate union of these two problems is a working communication.

Trying to navigate

Areas where there is a significantly problematic deep cultural assumption and presumption that these topics have been completely resolved whereas in reality they are very unresolved and yet taboo to talk about, resulting in predictable project and communication dysfunction:

- "types of systems": there is one! rational science!
- "explanations" there is one! rational science!
- "generalized stem" there is one! rational science!
- "what is statistics" there is one! rational science!

These have all been assumed to be 'affirmatively' resolved, whereas topics like ethics and mind and consciousness have been assumed to be 'negatively resolved':

- ethics: no, science has no room for superstition. we deal with hard truth!
- consciousness: "

- Books: the theory that would not die
- Escape from Model-land
- Systems thinking...something?

- Melanie Mitchel:
 - Artificial Intelligence
 - Complexity

- no-starch book?

background:

- we are all targets
- Cosmosapiens

characteristics of the space of mind vs. characteristics of the space of information-signal-networks. I am going to treat these the same, as 'mind' is less confusing and shorter.

Q: How much will quantum information theory change things?

chess-complexity:

- how did they know about turing chess
 - but not know he worked out minmax?
 - or were they playing for national credit (us not uk?)
 - and they credit shannon 1949 with minmax
-
- timeline inclusion or link in appendices...
 - - work in link to timeline

<https://en.wikipedia.org/wiki/Entscheidungsproblem>

Recusion

- - ~ debate between 'greed is good' zero-sum or negative-sum thinking vs. sustainable productivity.
- - You can't both argue the might wins by collapsing-right and that might is the foundation of right (or you can, but that's a self contradictory disinformation campaign which demonstrates the contradiction).

note look up very odd section on Francois Chollet on two hypotheses and statistical power...

- no null hypothesis?
- statistical power pre-test for sub-symbolic model?

4.4.10 realistic timelines: '36-year dev average'

4.4.# "proof of concept," vs. ready for market, the valley of death

- - ~ debate between 'greed is good' zero-sum or negative-sum thinking vs. sustainable productivity.
- - You can't both argue the might wins by collapsing-right and that might is the foundation of right (or you can, but that's a self contradictory disinformation campaign which demonstrates the contradiction).

'online voting' is likely an easy to understand common term for a more general and abstract process of "automated networked project decision coordination" with broad applications.

'AI' is itself a kind of interdisciplinary meta-area-of-areas: The different pathways and applications of AI are sufficiently different that people in different areas either don't know or just don't talk about each-other. Yet I would argue that it is (ironically) a useful 'general' abstraction to be able to talk about a vague ('general') 'big-tent of AI' that includes all such areas.

QUESTION: about 10min 20sec in, Eric Schmidt says that human user feedback 'does not scale' how so?

<https://www.youtube.com/watch?v=Sg3EchbCcA0>

"objective functions" in...what part of the system?

? Conclusion:

From the title we attempted to look at
'Definable Units of "Intelligence" for Evaluating AI Performance'
While "performance" should be defined in each specific context, some general recommended areas have been explored here:
- specification of skills
- specific tests
- areas of explainability
- ethics by intersection of areas
- ? explainability by intersection of areas
-

Introduction 2: Defining AI Goals

the people who want more 'explainability' do not define or keep changing the definition of what they are asking for. So, helping to define what explanation is being requested may be, if not emotionally satisfying, a substantive contribution.

What are parts of the overall picture of what people are looking for?

how system fail, reliability, and explainability:

-

<https://www.amazon.com/Artificial-Intelligence-audiobook/dp/B07YBHRX4B/>

<https://www.amazon.com/Audible-Deep-Learning-Python-Second/dp/B09RN70LT3/>

<https://www.amazon.com/Learning-Python-Second-Fran%C3%A7ois-Chollet/dp/1617296864/>

<https://www.amazon.com/Natural-Language-Processing-Action-Understanding/dp/B07X37578L/>

<https://www.amazon.com/Brief-History-Artificial-Intelligence-Where/dp/B088MMPZ49/>

As of end of March 2023, the time of writing:

- "ChatGPT Mar 14 Version" (not the pro version, just the older public version) is able to pass many of the previously unpassable AI performance tests, including Winograd Schemas (explain ambiguous pronouns based on world-knowledge), Sally-Anne Tasks (points of view of multiple persons), writing and analyzing computer code, basic (self inclusive) logical reasoning in a sentence (if a and b, then c).
- Many of the statements made about people (h.sapiens-humans), as a comparative frame of reference, are incorrect. In particular statements about the way h.sapiens-humans learn, statements about how h.sapiens-humans have responded to 'novel' situations historically, and how h.sapiens-humans even to present (2023) maintain policies on data, STEM, ethics, project management, etc., and awareness of human behavior (including the work of Khanaman & Tversky).
- Key concepts that really cannot be avoided in a discussion of AI such as 'generalization' are insufficiently explored and defined in most texts, often leading to undefinable, dubious, or self-contradictory statements.

Several of the lines of thinking that characterize the standard narrative about AI are not correct, or are no longer correct as of March 2023. And since the timeline of AI evolves so quickly, likely, hopefully, this document too will be out of date within a year. (See a timeline of publication dates and world events below).

As I have read through books about AI, books on low level nuts and bolts, and books speculating about the bigger picture, I started coming up with this concept of Object-Relationship-Spaces to clearly define what AI can NOT do, which was both the focus of some (not all) books and yet not clearly defined in any of the books I read.

To be clear here, there are now AI (or is one, Google Bard has been teased but I have not tested it or spoken with anyone who has access to Google Bard) who can

perform the tasks discussed in this section, and so the framing and perhaps semantics of the question need to be updated.

Introduction 2: Defining AI Goals

Navigating the World Before & After Large Language Models

This is a one-month paper, written between 2023.03.01 and 2023.03.31. These dates are significant, as this timespan is roughly the time period when many people, myself included, began to really be aware that Large Language Models in the form of OpenAI's ChatGPT had broken new ground somehow and had found a way to develop significantly beyond not only what previous AI models could do but also beyond what just recently published books said was likely or even possible. I started writing up notes on an Object-Relationship-Space AI-Operating-System perhaps a year or two earlier, based on gradually reading and rereading books including Melanie Mitchell's "Artificial Intelligence: A Guide for Thinking Humans" and Gary Marcus, & Ernest Davis' "Rebooting AI: Building Artificial Intelligence We Can Trust," then later Hobson Lane, et al's "Natural Language Processing in Action" and then Francois Chollet's "Deep Learning with Python, Second Edition 2nd Edition," and Michael Wooldridge's "A Brief History of Artificial Intelligence: What It Is, Where We Are, and Where We Are Going," just for fun while I worked a day job as a Data Scientist.

<https://www.amazon.com/Artificial-Intelligence-audiobook/dp/B07YBHRX4B/>

<https://www.amazon.com/Audible-Deep-Learning-Python-Second/dp/B09RN7OLT3/>

<https://www.amazon.com/Learning-Python-Second-Fran%C3%A7ois-Chollet/dp/1617296864/>

<https://www.amazon.com/Natural-Language-Processing-Action-Understanding/dp/B07X37578L/>

<https://www.amazon.com/Brief-History-Artificial-Intelligence-Where/dp/B088MMPZ49/>

I try to avoid internet and pop culture hype, so I did not look into ChatGPT right away. In fact, I probably did not start testing ChatGPT until halfway through the month of March (which is also when they released the March 14th update, and was before the Pro version). And whereas I had been, I had started, writing a paper, based on all these books, and tests I was running on other AI such as Wolfram Alpha, with a focus on more clearly explaining what AI was not able to do and what more specifically we want AI to be able to do. I was motivated to take this approach, to try to create a framework, because the explanation shared by all the books that AI was unable to ever do 'stuff' because it did not have special human abilities, which I knew very clearly from my past occupation as teacher and educational program developer were very much NOT accurate descriptions of human learning and innate ability, seemed vague pessimistic and impractical. AI was stuck at a given level, and I was confident that we could better understand what we wanted to do and find a way to do it if we took a different approach to specifying the details. Then the whole world turned upside down. All of sudden, there is AI that can do pretty much all the things that all my books said were impossible. But there still was not (in

general circulation) a good way to describe what exactly chatGPT could do that other's could not, what it still struggled with, and same as before where we wanted to take AI system design further in the future.

The specific topic of this paper had not changed at all, and most of the sections were the same. But this part of the paper, on how to think about books on AI, just became much more convoluted to explain, the wording of 'what AI can't do now' had all just changed or dropped off the map, and the imperative to focus on future development and uses of the framework given that Large Language Models had lept the game ahead into the future by bounds, had gone from: "I am crazy to try to push forward into the future with this," to -> "I would be crazy NOT to try to push forward into the future with this."

I was kind of stuck on this see-saw, where half the world was in a bazaar past where people gave vague explanations for why strong AI was impossible (and yes I had read Kurzweil decades before, and no I do not think he is wrong about everything he predicts), and then the future half of the world was giving equally inadequate vague explanations, and now greater fears, about AI being able to do everything.

So I apologize for the bazaar nature of this section, where some of the comments and analysis may simply seem 'out of date.' But I really want to emphasize the need, though it looks different in the world before ChatGPT and in the world after, to clarify what we are saying and to not fall into using frames of reference that may be meaningless.

1. In the past the context was: No AI can now nor likely will ever be able to do X (where X is often not very clearly defined, even in the case of test questions, for example there are very serious problems with some Winograd Schema questions including the often-used 'first' question).
2. As of March 2023 some of the same ability questions can and should be discussed in a well defined way, but the context needs to shift to clarify what specific AI components can or cannot perform a Task and whether that skill is relevant to what the project-role of that AI is. For example, does it make sense to require an image-classification AI-component to also answer natural-language questions about images? In some cases the answers will be yes, in some cases not. (Whereas in the past the inability was absolute so the component-role-project distinctions and context never needed to be clarified or even brought up.)

For example, we can ask questions and see what specific things, what specific objects and types of objects, the AI is, or is not, able to deal with. Is the AI able to recognize these objects that we are asking it about? Is it able to see how some objects relate to other objects ("relationships")? Is the AI able to combine objects? Etc.

The idea is using objects, types of objects, and relationships between types of objects, as a framework for defining AI skills and abilities. And if you hit a wall where you really want AI to do something, but it can't do it, then you could figure out how to do it. For example, make a general AI operating system that breaks things down into objects and basically has a database of objects that it is dealing with in that given context, not a knowledge-base about the whole universe, but objects that are specific to the project you are trying to do and the skill-goal you have set (your target): an external project-object-database.

Separating what AI is extremely good at and extremely bad at seemed to help identify low hanging fruit; and the original paper started around the phrase, "AI is very good at identifying things, but it's not very good at doing anything with that information."

For example:

You can show an AI two pictures and it can correctly identify exactly what sub-species of animal those are pictures of, but then ask the AI something about the pictures like: Did I just show you two pictures? And it either can't answer correctly or can't answer at all.

Released in late 2022 and then upgraded in early 2023, a 'new' chat-interface AI called 'ChatGPT' was released by OpenAI. As of march 2023, ChatGPT can do many of the 'object handling' tasks that most books on AI both spent most of their text 1. explaining in their view how and why cannot now do these tasks, and 2. how and why AI may never be able to do these tasks. Usually in the same circular narrative: general-human, etc.

But as of march 2023 we have an AI that can 'do things' but we still lack a good framework for describing what AI could not do, what most AI cannot do, what ChatGPT can do, and more broadly what we want various AI to do in future in specific contexts.

Part of why I took that line out of the paper is that as of march 2023 OpenAI's ChatGPT (released in late 2022 and since upgraded regularly) is so much better at parts of these tasks of dealing with objects, and even doing some in a very flexible and open-ended way, that it is no longer factually correct to say "AI cannot deal with objects in general." We now have one that can.

And I have never seen 'handling objects' discussed in any articles about ChatGPT, even by OpenAI, or about any AI. Most news articles (like this [one](#):) unsurprisingly say exactly the same standard line: ChatGPT isn't really intelligent because it lacks understanding, it doesn't have

general-intelligence because it isn't like people, and on and on in that same cycle leading nowhere.

Interestingly, ChatGPT cannot deal with pictures, it strictly deals with text. So whether it could connect a picture to text, we don't know. But amazingly, ChatGPT not only has the logical-problem-space-skills we were just talking about, it also has a fair amount of common-sense world-knowledge-base information.

One of the things that is important for this discussion of how we should go about designing AI is that arguably my crude idea of having a brute force project-object-database is too-brute-force for many tasks. We would want the AI do a lot of things on its own. The AI wouldn't need to analytically separate all the objects and step through every step the way people might lay out a process workflow. There is a long history of approaching some AI problems this way often called 'symbolic' AI, including expert systems (some of which do work well) using hand-crafted logical steps for everything. We tried for many years and it does not work well as a general method for most problems (but it is ok for some).

However, aside from the information processing that goes on inside the AI (e.g. how it identifies the picture as a cat), if we are talking about a **project**, if you have an AI that is just one of many participants working on a **project** then it needs to be able to communicate and coordinate about everything that is in an external project-object-database with all the other people and other AI who are working on that project.

And in some very specific cases, because much AI is micro-specific AI not super-grand mega-AI, that if you had for example a local AI running on a tiny resource limited micro-controler computer, maybe in a factory or warehouse somewhere with no internet access, no big-computer access, and it was having trouble dealing with an object in a very specific task, then why not try having it use a project-object database to help it carry out a specific process.

This has wandered off the topic of books that I would like to read. I think it would be really neat if a book talked about possible ideas in the future:

- How is AI going to integrate with Biology?
- How does AI relate to education, governance, legal systems, STEM & Science?
- How will AI be directly and indirectly integrated into the human body?
- How might AI be used during space programs?

And general mind discussion. You have the human mind, you have, theoretically, Extra Terrestrial species that have minds, and then AI-mindspace. And then you have all the other mind-space topics that we are not supposed to talk about: sacred plants, afterlife experiences, meditation, mindfulness, compassion, ethics, boy scout values: duty, responsibility, conscience. I think we should talk about those things. I would love to read about those things.

How universal is mindsphere? There is only one species with language and self-reflecting consciousness on earth. Is every intelligent (Drake Equation) species going to have some behavioral problems similar to those of humans? How many of the weird problems that people have are just part of universal mindsphere? Mindspace is difficult, for example the 'curse of dimensionality.' Dealing with higher and lower dimensions is probably difficult for anyone anywhere, not just for some people or species or organizations. As we build up our general model of mindsphere, what are the things that we think are universal? What are the things we think are not universal, and which are things we just don't know? This will be important both for AI and just managing societies and facilitating human development.

What I would put into that discussion are things like general system collapse, and categories of types of systems. I suspect there are a small number of things that are very universal, as in probably in any universe: some aspects of STEM; some aspects of categories of types of systems; and general system and definition collapse processes. Even if that is the only skeleton of universality that would allow us to communicate with other species or even be the best way to interact with AI if it thinks very differently from how we do, then that may be the low hanging fruit that we can start with.

And hopefully the better we understand the basics of mindsphere, the better foundation we will have in many areas including properly managing AI.

There is a short and long recommended reading list here. The goal here is not to summarize what is in these books. The hope is that you will read these books for yourself (listening to the short list on audible would take just a few days, not unreasonable). The goal here is to equip you to pragmatically use the information that you will find out in the world, whether using these books or other. To prepare you for issues, to prepare you to jump further, to prepare you to ask the hard questions and to make concrete plans to find practical solutions.

A Short List:

The framework list that you will see in part one may be overwhelming at first to imagine as a giant chart or grid or array or matrix of layers of intersecting areas. But we can start with a smaller list. You could even start by making your own list of the things you are most interested in. Maybe you are mostly interested in time and time series and schedules and histories. Maybe you are most interested in projections and predictions. Maybe you are interested in physical spaces and physical interaction. Maybe you are interested in natural language. Maybe you are interested in 'gamification' simplified game rules. Maybe you are interested in comparing hypotheticals in forensic reconstruction, or in defensive preparations. Whatever spaces of events and interactions you are most familiar with, start there.

(Review)

AI must be able to handle **objects** in the following inter-related contexts:

1. object-relationship-spaces
2. project-object-database (in project-framework)
3. project-participants (in project-framework)

Such that these are defined as existing outside of the AI, and that so long as the AI effectively deals with these, it does not matter whether how the AI does so

(Note: in GEB, D. Hofestedter made the perhaps counter-intuitive prediction that AI might not be able to count quickly, which might haven been an insight into the quagmire of balancing system-1 and system-2 type processes.

Suffice it to say, we should stay flexible and not lock AI into one way of solving a problem.

note:

biological systems...

To reiterate:

AI must be able to handle objects in the following inter-related contexts:

1. object-relationship-spaces
2. project-object-database (in project-framework)
3. project-participants (in project-framework)

such that these are defined as existing outside of the AI, and so long as the AI effectively deals with these it does not matter how the AI does so.

A hopefully nuanced note on disagreement on details:

As much as possible I would like to applaud and praise and advocate for Francois Chollet, his Keras, and his book Deep Learning in Python 2nd Edition (I have not read the first edition), while also disagreeing on specific points. I do not want my use of his excellent instructive explanations of the current view to be misinterpreted as criticisms of Francois Chollet himself in any way. This may be at times an impossible needle to thread, but I would like to disagree on certain points without being disagreeable personally.

For these questions we are not asking the AI to know anything about the world at large or even about cats at large. We are asking about basic operational data that would be in any project-log of what has happened. We are asking tautological questions about that information. And that information is information that the AI already has.

- Two questions were asked.
- Questions were answered.

- The answers are the same.

And there are two, one of them somewhat comical, ways in which the AI does not have the ability to do this.

1. Assuming this is an AI that has been trained for some language function, we have never given it (or trained it to give itself) an 'object space.' It has no problem-space or project-space to use the information that it just gave you. It literally has the answer. It just gave you the answer. But it can't project-manage that same correct answer as an object in object-space. It has no memory, it has no internal model (of whatever kind) of what it has itself produced with a granularity of objects covering what you are asking about.

2. (the comical reason) If it is a picture-only AI then it can't process language questions, period. It has no way of processing anything in language at all. The input is one picture. The output is one word. It wasn't built to do anything else. It's like asking a screw driver a question about what you just used it for. This might sound trivially obvious, but most AI historically (and often for practical reasons) have been this kind of hyper-focused single-purpose AI that has one very narrow type of input and just one very narrow output, like cat-picture classification. And yet people vaguely criticize it for being what it is: one single-purpose component that works rather well. In some cases it is a problem that the AI system as a whole cannot do something, but often people do not have any practical or clear goal in rambling on about the ineffable lack of stuff-ness and you-know-what about the AI component. And there are very good product-design reasons why having an AI do only what it needs to do, and no more, is a very good idea.

If you show a picture-identifying-AI two pictures, the only things that exist in that AI's world, in the problem space, are two things: the two answers it just gave you. If you want to ask the AI if those two answers are the same (both 'cat'), not only does the AI have the answer, the only things it has are the answer. If that is your goal then the best, most efficient, route to getting the AI to answer just that logical question about the answers it gave you is not to give the AI infinitely more "knowledge base" and "common sense" knowledge about the world at large: the AI already has the answer. The problem is that the AI cannot deal with objects. The ability to deal internally (for itself) or externally (like in a group project) with those object does not depend on a huge knowledge-base-graph, which would bloat the AI enormously.

How many things are in your problem space. -> Two.

Are they the same? -> Yes.

In theory, ChatGPT may have achieved a kind of 'general-knowledge' + object relationships AI, and for the sake of argument let's say it relies upon deep-learning higher order concept embedding vectors. (As far as I know there is no knowledge graph in chatGPT). Depending on the needs of your system, if you need a local efficient mechanism, needing to connect to a vast

international network of specialized sub-servers and waiting a while for the answer might not be a viable option.

In many discussions, very different goals for AI get all mixed together, and people end up saying things that are just not relevant to the goals in question.

- Generalization
- Knowledge base
- Transference of learning
- general criticisms of 'lists of numbers'

None of these have anything to do with the simple fact that the AI was not given the specific ability that we are testing it for: The ability to do something very specific and well defined with information that it already has.

Does the system need that ability?

Are there any reasons why the system should not have that ability?

Note: we do not know specifically or generally what can create this ability. Some writers insist they know the one-needed-approach (such as 'Rebooting AI').

We don't know if there is an 'only one' direct mechanism of human understanding.

Two examples, in Rebooting AI: the whole book is about there being only-one solution to giving AI the ability to deal with objects (at least I think that is what they are talking about...they are not entirely specific) and that single solution is what they call 'classical AI' and what the standard jargon calls 'symbolic AI' (not a good clear name, but that's the official jargon) meaning hand-crafted (by people) sets of rules and steps for 'reasoning' the way people think that people 'reason.' Then in the conclusion, his first 'advice' for the reader is that there is no silver bullet, no single answer. AI will be as diverse as biology with lots of solutions to lots of problems. I ask: How can you have both 'only one' and 'not only one' in this situation? Maybe I am too Aristotelian in my thinking.

And again, in Francois Challet's 2nd Edition, chapter 14: He gives several different 'only one' explanations for "human general intelligence." Again, my question: How can you have multiple different 'only one' solutions?

We do not know what technologies are going to do well at what tasks without empirically trying and comparing them. And we do not know what specific tests will be similar results to what other tests, in other words: a technology that is best for one kind of problem may not be best for another kind of specific-situation-problem, regardless of whether we think those problems 'should be' the same or should be different.

A stock clickbait title for AI papers has become: "The unreasonable effectiveness of ____ method." Because we so often cannot understand why the technology works so well.

Let's be honest about what we don't know yet, and try to be as clear as we can about what we are trying to do. The best I can do for this task is the object-relationship-space framework + project-context.

1. Descriptions of AI in the Literature

A very common statement in many books on AI goes like this:

*AI can identify a picture as a cat,
but it doesn't really "understand" what a cat is.*

There will be more discussion of this discussion in Part 3 of this paper, but for now, let's think about a few details of this statement.

Let's rephrase the statement slightly:

A single-purpose picture-recognition-AI, does not 'really understand' the natural-language word "Cat."

Now, let's set "really understand" aside for now.

Let's re-rephrase the statement:

A single-purpose picture-recognition-AI, does not _____ the natural-language word "Cat."

What would it make sense to put in that blank?

'have the ability to process'
'handle-as-data'
'train on'

Those would at least make sense. But let's put 'really understand' back in and look again:

A single-purpose picture-recognition-AI, does not 'really understand' the natural-language word "Cat."

A natural language component is, often, separate from a single-purpose

Unknowns and Unknown-Parallels

- Will there be portable mostly-internal systems?
- How does the human brain work?

Details present or absent in discussion: Do we know where we are, and where we are trying to go? How do we get there?

AI OS & Architecture Problem-Space(s): Design to Solve

	<u>internal</u> project objects	<u>external</u> project objects
<u>single</u> component		
<u>networked</u> components		

<i>types of signal/ data/ types of systems [NLP, images, sounds, etc.]</i>	specific-deployment (portable)	general (less portable)

Kasparov said ~Chess computer's don't get tired"

Well, not if you leave them plugged into an infinite power supply grid supplied by an entire country and paid for by dozens of other organizations, but that's not exactly a realistic use-case for anything in reality. Though the point of 'showing tiredness' as a h.sapiens-human would be certainly a good point. If the machine did have to throttle itself back, and you could hear how fast the fan was going, and feel how the cooling system was getting, would that be similar on some ways? And is it fair to compare a 'machine' that happens to be in a distant secret room with a person who you see in front of you. For example, you could play blind-proxy chess, where the chess-players are in secret rooms surrounded by support teams and the players at the board are merely opaque dummy terminals that can't see the other player. It kind raises the 'on-edge' design and deployment question, which in AI-ML jargon means a system that really deployed on site with realistic resource limitations, not something running in the cloud.

If I can add one more 'dimension' to this thread of thoughts:

A common topic is 'higher dimensions' and 'high dimensionality'

If the problem space (possibly a loose use of the term) here in our robot design is the 18 factors mentioned above, if we were making 'machine learning' or 'ai' model to make predictions on a dataset with many data points about many azzi-robots with information in all these 18 areas, we would call these areas 'dimensions' in the model.

One of the reasons why I bring this issue of 'high dimensionality' up is the very common notion that people cannot understand any system with more than 2 or 3 dimensions, because we can only visualize and graph up to 3 dimensions. But I would caution skepticism on this point. Yes, a 10,000 dimensional space is probably quite hard to understand in one-go, but I think the above rough list

of items should give you a sense that you can indeed understand the inter-relationships between even about twenty factors, even though you could not express all those relationships easily with a single line on a two dimensional graph.

Sometimes the issue of parts and layers may more abstractly extend inward

Note: This also may raise the perhaps somewhat fractal recursive design question of having 'brain' layers in the network to manage lower levels of components. Perhaps echoing the human 'homunculus problem' in ~psychology of 'finding the brain part of the brain that controls what the brain is thinking about and doing,' which of course is an impossible problem (there is no brain in brain, or a tiny person in a tiny office in the brain, or a tiny person behind the reigns of a chariot in the brain as the greeks put it). Yet for AI, there likely will be 'brain layers' that help to organize larger networks.

- sound coding and AI

todo:

audio from 2023.04.2

SP? Cordwainer smith?

Phrases like 'solve AI'
or 'solve intelligence'

and not Education-ology/Learning-ology field concept,