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.

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

Table of Contents: Detailed

```
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 tallent, could, like Sherlock Holms, 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 Kasparaov 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-simply 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 Kasperov 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 unprecedented 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)
            - develope, 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: <u>before</u> 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: <u>after</u> 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=rnIgnS8Susg
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 wearhouse AI 'participation', or if, even after the ChatGPT revolution, there is no intersection, interconnection, and the only things that AI can do in a wearhouse 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:

> "How many pictures did I just show you?"   [Two]
> "Did I show you a picture of a cat?"   [Yes]
> "Were the answers you gave me both the same?"   [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 scenareos 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 wearhouse 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
manageably 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') understanding 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-tranfer-learning-intelligence-concisouness special sause 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 'lingustic 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 ot 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 doublt 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 partiipans in a projects-space? How will image and language concpets work together across projects-spaces using dignital and AI tools? These are practidal 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. 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 a 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

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]	C. The C is for computer. *Note: She will also give this answer if you replace C in the question with D. Apparently she likes alphabet*	B + B = C	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. In mathematical notation, the transitive property of equality can be expressed as follows: If A = B and B = C, then A = C. 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. Therefore, A = C = 2.

	games. Adorable.		

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.
		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
		
		Note: Wolfram Alpha can generate math graphs and charts.
			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.
			
			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.
			
			Notes:
			A: This is a correct understanding of the question.
			B: ChatGPT is Text only. No the ability to show pictures, graphs, etc.
			
			But!...
			
			Q: Can you make ascii art?
			
			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:
			
			`  _____`
			` /      \`
			`|  ()  () |`
			`|    ^    |`
			` _____/`
			
			Please let me know if you have any specific ASCII art in mind that you would like me to create!

<table>
<tr><td colspan="4">

Note: The AI can either retrienve ASCII art from a search, or generate a new one.

Q: Please generate an ascii art picture of a sun and show it in a markdown box.
A: Certainly, here's an ASCII art picture of a sun shown in a markdown box:

```
 \  /
  .-.
-(   )-
  '-'
 /  \
```

</td></tr>
</table>

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 could design and 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 being 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 evaulation 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 travers 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-loose-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 an 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 Imanual 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 mindspace?

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*
> <u>Interpolation</u> *can only help you make sense of things that are very close to what you've seen before: it enables <u>local</u> generalization. But remarkably, humans deal with <u>extreme</u> <u>novelty</u> <u>all the time</u>, and they do just <u>fine</u>. You <u>don't need to be trained in advance</u> on countless examples of every situation you'll ever have to encounter. <u>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</u>. 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 <u>extreme generalization</u>, which is enabled by <u>cognitive mechanisms</u> other than interpolation: <u>abstraction</u>, <u>symbolic models of the world</u>, <u>reasoning</u>, logic, <u>common sense</u>, <u>innate priors</u> about the world--what we generally call <u>reason</u>, as opposed to <u>intuition</u> and <u>pattern recognition</u>. The latter are largely <u>interpolative</u> in nature, but the former isn't. Both are essential to <u>intelligence</u>. We'll talk more about this in <u>chapter 14</u>.*

h.sapiens do <u>not</u>:
- 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 <u>do</u>:
        - 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 that 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* <u>*optimization*</u> *and* <u>*generalization*</u>*. The ideal model is one that stands right at the* **border** *between* <u>*underfitting*</u> *and* <u>*overfitting*</u>*, between* <u>*undercapacity*</u> *and* <u>*overcapacity*</u>*. 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"
     - "Generalication"
     - "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 genetec 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 generalie
      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
realness and treat it as though it exists, though it may either no 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 the 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 lake 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 i 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-phyisical 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 Fransoi 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 centry 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 Hofstedter 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 jubblies 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 the 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 trajically/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, 'Choas' & '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
                - Sigificant 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 Influences on Model Architecture
                9.3.1 chollet:
                - peer review breakdown
                - taking out parts of models
                - oblation studies?

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, mabe 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 benifit 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 awave 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 an 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:

(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

Here is a sample 'we can' agenda statement:
(Which is more or less the same as the agenda statement for definition behavior general system collapse studies.)

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 Definition-Clarification vs. Disinformation Attacks
                - Generalized Categories of Types of Systems
                - Generalized Ethics, Duty & Responsibility
                   value-function-meaning, boyscout-project-values,
                    Duty & Responsibility

                - (maybe) Generalized (Sustainable) Productivity
                - (maybe) Generalized definition behaviors.

(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
             - 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 knowns) 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 is 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 'trolly problem' (regardless of that scenareo'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 frequency
reiterates, intersection/interlocking areas of STEM, project-context, boy-scout
values (or perhaps now just 'scout' values)

      STEM, Ethics & Mindfulness

      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 question-space

4.3.15 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?

4.3.16 Analogies?

4.3.17 system epidemiology

4.3.18 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.

4.3.19 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.

4.3.20 '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/11DFQtsNjrqHENS0D7UpuZhOhcqCKK39JfmEBc8O8NHI/

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