
vos.ai
agi

We thank you for your consideration and welcome the opportunity to embark on this journey together for a better future.

CONTENTS

1. Problem Statement

This section describes the opportunity and context for the problem being solved. It will clearly articulate how the problem will be solved using AI.

1.1 Problem

Our history is filled with countless innovations, innovations that have had unpredictable side effects on our world. Many, if not all of these have impacted us both positively and negatively – whether it be socially, emotionally, physically or spiritually. For example, the smart phone is as common as a wallet, yet people are more concerned with their social media updates rather than their surroundings. In extreme cases, crashing their cars killing others on the road.

Within the next decade, we expect innovations in automation unparalleled to anything seen before. Drones, robotics, driverless cars, and artificial intelligence are just a few examples.

The problem we see is within the invention itself. The invention of AI, specifically Artificial General Intelligence (AGI), will be one of the greatest challenges we have ever faced. It will disrupt governments, economies, workforces, social, physical and emotional aspects of our lives. For the sake of clarity, an Artificial General Intelligence is the intelligence of a machine that could successfully perform any intellectual task that a human being can. It is a primary goal of some artificial intelligence research and a common topic in science fiction and future studies.

What will the workforces of tomorrow be like? How many jobs will be lost or simply non-existent? Will our kids have careers like we do today? What kind of social interaction will people have? Even worse, will AGI be used as a weapon?

What's most sobering is the fact that we have no defense for our workforce and society in general from threats rooting in AGI.

<u>The problem:</u>

"How is an AGI created and incorporated into daily life that only allows the beneficial aspects and excludes the negative side effects?"

1.2 Opportunity

We have an opportunity to create an AGI capable of functionally learning as we understand humans do. Specifically, how we learn and grow as humans, from the smallest of ideas to the grandest of concepts.

We can embark upon this opportunity with our eyes wide open analyzing the positive and negative impacts of this technology, while properly assessing how to incorporate it into our lives. This allows time to adapt to the coming technology before it overwhelms us with the powerful computing power just on the horizon.

Lastly, we capture what it means to make a true AGI before it gets ahead of us. Capturing this insight allows us to properly set forth architectures, rules and guidelines to ensure these free "thinking" AGIs do not run wild causing damage outside of our control. In other words – a proper architecture for building AGI ensuring benefits over negative effects.

1.3 How AI applies?

At the very heart of this plan is the creation of an AGI. The creation of an AGI includes leveraging modern day techniques such as artificial neural networks, genetic algorithms, knowledge bases, hybrid data stores and CPU/GPU processing power in a distributed messaging architecture conducive to scaling infinitely.

The AGI created will go through multiple iterations each building upon the previous steps. We will analyze the results of the AGI and its effectiveness through multiple public outlets including but not limited to social media, email, chatting, media websites and the like. From this effort, we will properly gauge the accuracy of our approach and perform validation at the most basic of levels.

Lastly, the end result is the seamless incorporation into existing workflows in everyday life, working symbiotically with people in an effort to aid in their efforts with communications, knowledge acquisition or related functions. This symbiotic relationship enables the AGI with the capability to learn over time cooperatively with humans while allowing the human aspect to naturally train and feed the AGI information required for learning. Overtime, the AGI adds value to human counterparts for improving overall efficiencies in the arbitrary tasks at hand.

1.4 Why is it timely and important?

The tipping point for innovations derived from AI is upon us. Organizations around the world are constantly pushing the limits of innovation from startups to Fortune 100s, and universities to research institutions. They are addressing core business cases, exploring basics of the technology or creating new products.

These innovations are catching everyone off guard with little or no thought of the downstream effects. Those familiar with these techniques are looking to the technology to replace people in the workforce. For example, the AEC (architecture, engineering, construction) industry is leveraging computer vision and machine learning to help reduce and / or ultimately remove the need for surveyors and inspectors. Automating this process would improve the efficiency of their work while reducing costs. Yet, there is no thought as to what will happen to these people being replaced let alone the adverse effects.

It is crucial for the future of AGI and people as a whole to incorporate AI and related technologies in order to compete with the future demands of life where innovation occurs faster than we can keep up. Without proper incorporation, we can see adverse effects that could impact the AGI directly due to lack of sufficient and/or accurate training. Moreover, organizations leveraging AGI would begin to alter operations such that we remove humans from the equation. At first, these organizations may experience immediate gains in both efficiencies and costs, but over time, they may experience greater problems due to inadequate or improper training of the AGI. This could ultimately lead to the demise of their very own organization.

1.5 Addressing the problem

The solution to this problem applies across the board to multiple domains. **Any process today requiring human interaction is a candidate for automation.**

By addressing this problem, we are able to identify and categorize key aspects subject to automation. Identifying these aspects or functions provides the patterns and means to gradually introduce AI to enhance and augment the functions rather than completely replacing them. Ultimately, we foresee that we can augment humans with an AGI to improve the quality of work and life, provided it's integrated properly.

2. Proposed Solution

This section explains the approach used to solve the problem defined in Section 1 in detail.

2.1 Overall approach

The overall approach is encompassed into the four phases below.

2.1.1 AGI

The creation of a AGI leveraging commonly known techniques which include artificial neural networks, genetic algorithms, knowledge bases, hybrid data stores and CPU/ GPU processing power in a distributed messaging architecture conducive to scaling infinitely.

The AGI learns patterns over bytes at its most basic of function. For the initial purposes we will direct this towards images and written languages, specifically the English language. The key differentiator to this approach is that rather than focusing one or a handful of ML/CV algorithms against a specific dataset, we are applying a particular recipe of algorithms to a broad set of data sets such that it's generic enough to be applicable to all while not narrowing our focus on just one specific context.

The proprietary methods of the AGI is a combination of these techniques in a pattern that is analogous to the assumed workflows of the human mind and how it learns.

The deliverable for this phase is a bare bones AGI software application with limited training and user interfaces.

2.1.2 Initial Training

The purpose of this phase is to seed the training used for learning. Training will comprise of known English words, books, stories and isolated data created by human interactions. The isolated data created by human interactions may be gathered manually through a crowd sourced effort or by means of transcripts of available conversations either within the public domain or within our own private data collection processes.

The deliverable for this phase is an evolved AGI based on the efforts from section 2.1.1 combined with the training and smoke testing of the system.

2.1.3 Validation

The purpose of this phase is to expose the AGI in a controlled manor to public outlets (e.g. social media, email, and chat) in order to validate its ability to learn language over time without the need for commonly used NLP techniques. This validation will be monitored at all times and evaluated both externally without interaction as well as during interaction by selected team members and crowd sourced efforts.

The deliverable for this phase is a validated and evolved AGI based on the combined efforts from 2.1.1-2.1.3. It is theorized at this point the AGI is mature enough to converse free flowingly either by listening and reacting to stimulus or prompting its own stimulus in hopes of a reaction.

2.1.4 Identification

The purpose of this phase is to identify key functions in existing work life that could benefit from this technology by having a symbiotic relationship. Identified candidate cases are fielded from numerous sectors ranging from common jobs to the most complicated jobs.

The deliverable for this phase is a set of identified cases and how the incorporation of the AGI will take place within functions for work life for selected candidates.

2.1.5 Incorporation

During this phase, we will produce various candidate applications which lend towards the integration into normal work life for selected cases from section 2.1.4. We may leverage organizations interested in this concept and willing to cooperate with the effort.

The deliverable for this phase may either be one or more prototypical applications demonstrating the seamless integration into the workforce working collaboratively with human counterparts. Initially the AGI learns from data generated by the human counterpart performing their daily duties. The end goal is to have the AGI begin to enhance the work of the counterpart however possible when related to communications in the English language.

Alternatively, we may opt to integrate with previous integrations from section 2.1.3 or new integrations providing the capability to reduce the scope of work and focus more so on the AGI.

We do expect this portion of the project to slightly mutate depending on our findings over time from previous sections.

2.2 Technological approaches

2.2.1 Hardware

The creation of AGI requires specific underlying hardware which is not necessarily common to the world today. We plan to leverage hardware which may include GPU and / or FPGA components in order to perform operations faster and more efficiently than a CPU.

The AGI requires the fastest available processing. Therefore, we may or may not use a cloud-based solution (e.g. AWS, GCP) due to the bottlenecks and performance issues arising specifically with virtualization and NVidia GPUs.

There are 4 categorizations to the components of the AGI that require specific hardware and performance.

- CPUs
- GPUs / DPUs
- FPGA
- Data stores

2.2.1.1 CPU

When it comes to CPUs we feel that either Intel or AMD based CPUs should suffice for any component requiring a CPU. We do foresee the need for many cores and threads as well as multiple CPUs per server over a cluster of servers.

2.2.1.2 GPU / DPU

In regards to GPUs, there are particular nuances between the different manufacturers of GPUs. Specifically, between AMD and NVidia. Primarily, when it comes to how they function in a virtualized environment. The AMD based GPUs are the only ones that truly guarantee a virtual system gets dedicated time to use the GPU whereas NVidia does not.

Moreover, when it comes to costs at scale, AMD based solutions are far less in costs and power consumption when compared to NVidia.

Our objective with the GPU aspects is to remain as agnostic as possible by not developing a solution that is tied to one manufacturer or the other. This will further be explained in the next section. The developed AGI will be capable of adapting to its underlying hardware. It is assumed any server where the AGI resides will have at least one (1) GPU while taking into account it may be

distributing its processing across multiple GPUs or a cluster of servers with multiple GPUs.

We foresee the use of custom servers created specifically for the purposes of training and learning as well as cloud-based solutions to properly address the requirements and demands of the AGI.

In regards to DPUs, this is a fairly recent technology surfacing which alters the standard architecture of modern day computer systems. A DPU based system is designed solely for machine learning algorithms where processors are integrated and architected in a pattern best suited for the data volume and intensity of the ML algorithms.

2.2.1.3 FPGA

We have yet to rule out the need for an FPGA backed solution for the AGI. We will explore this avenue for its viability and practicality. The main objective with FPGA is to further improve the performance of highly specialized algorithms that may not run nearly as fast on a GPU.

We also consider alternative means of computing machine learning algorithms in this section which may or may not be FPGA based.

2.2.1.4 Data stores

Normally one would think that a standard relational database would suffice for an AGI. It could very well serve a proper purpose. However, we foresee that non-standard data stores are required for a proper AGI. For example, the use of document, key-value and graph databases will be employed for this AGI.

These data stores will grow to a very large size and will likely scale to many servers. Therefore, the need for adequate storage space will be required. If we find throughout the course that the addition of meta-data (e.g. images, audio, video) to the data stores is required this may increase the amount of storage required.

If the underlying hard drives are spindle-based they should spin at no less than 7200RPM. Solid state storage is preferred.

To summarize, we theorize the following hardware specifications suffices.

Basic Processing (CPU)

- 2-4 servers load balanced
- Single or dual multi-core CPUs
- 8 GB RAM minimum
- 64-128GB Storage (Spindle)
- 1x Gigabit connectivity

Core Processing (CPU)

- 2-8 servers distributed
- Dual multi-core CPUs / server
- 16 GB RAM minimum
- 512GB-1TB Storage (Spindle)
- 2x Gigabit connectivity

Advanced Processing (GPU)

- 2-8 servers distributed
- Dual multi-core CPUs / server
- 32 GB RAM minimum
- 1TB-5TB Storage (SSD)
- 4x Gigabit connectivity
- 8-way (minimum) GPU
- FPGA requirements TBD

Data Stores

- 2-4 servers clustered
- Dual multi-core CPUs / server
- 64 GB RAM minimum
- 10-50TB Storage (SSD)
- 4x Gigabit connectivity

These specifications are intended for servers only. Development workstations and servers will be comprised of the absolute bare minimum in order to develop properly.

2.2.2 Software

2.2.2.1 Basics

We intend to leverage open source software as much as possible. Therefore, the operating system(s) of choice will likely be Ubuntu for workstations and Debian for servers. Alternatively, we may choose Red hat or Cent OS. We do not foresee the need for Windows or Apple based operating systems.

2.2.2.2 Languages and formats

We foresee the use of various programming languages.

- Python
- C / C++
- Assembly or alternate lower level languages (if required)

We will likely leverage JSON or XML formats for messages and / or configurations. Binary serialization may also be employed where required.

2.2.2.3 Best Practices

For any development performed we will adhere to open standards conducive to creating proper software solutions. For example, Python has a style guide (PEP8) which will be strictly adhered to throughout the development process.

Documentation of all code, architecture, designs will be a part of the process of constructing the AGI.

We will likely follow a continuous integration process for deployments as well as managing builds between environments. The management of builds may include code analysis, code testing and deployments between environments.

Lastly, all source code will be controlled in a source control system (e.g. GIT).

2.2.2.4 Architecture

The AGI will be developed following best practices for enterprise software. We foresee a highly distributed messaging architecture where state is never stored within any component and all components share a canonical messaging format. The architecture will allow to infinitely scale according to its underlying infrastructure.

Little to no thought will be given to encrypting internal messages at this time. However, securing this solution in the future will be very simple given that all best practices for building enterprise software will be adhered to.

2.2.2.5 Data stores

The use of hybrid data stores (e.g. NoSQL) will be employed as the primary data store back end for the AGI. We intend to use Graph, Document and Key/Value data stores. There may be two options for a data store.

The first option may be a multi-modal database possessing the requirements for the AGI. The latter may be three different data stores. One for each type of store. This will be determined at a later date.

Regardless of option chosen, our intent is to ensure the solution can meet the demands of the AGI. Moreover, the option chosen must be able to scale with the system automatically with little to no performance degradation.

2.2.2.6 3rd party components

Any software system requires the use of integrated development environments, tools and libraries to accomplish their objectives. This will be treated on a case by case basis.

2.2.2.7 Machine Learning

As stated in section 2.2.2.6, we will leverage 3rd party components on an as need basis. Our intent is to combine existing frameworks for machine learning with our own proprietary constructs and patterns to support the AGI development.

Any libraries chosen leveraging CPUs, GPUs, DPUs, or FPGAs will be chosen such that they are agnostic to the make, model and manufacturer of these hardware components. This effort will be made at every step of the way wherever possible. There may be instances where this cannot be done. However, our goal is to remain agnostic.

For example, rather than leveraging CUDA libraries for NVidia the use of Open CL may be employed. Open CL would allow for any GPU to be utilized regardless if it is from AMD or NVidia

2.3 Exponential impact

The chosen architecture and design of this AGI starts with the core concept of our design. Our design primarily works at the byte level. Sequences of bytes are then categorized accordingly. Without categorization, the AGI would be unable to determine if the bytes are a video, image, sound or text. Training and learning are assigned to these sequences of bytes. The initial approach will cover spoken languages from any origin with our focus on the English language. However, the same approach with no change to the system allows for the inclusion of images, video and audio I/O. The system will be capable of "re-learning" information and assigning new artifacts.

For example, the AGI learns the word "dog" and understands a few things about a dog at the language level. Thereafter, we could show it an image, video or sound clip of a dog. The new data would then be associated with the dog. The AGI would be able to consume as well as produce these images of its own accord. This is merely an example of the future impact. The production of images may or may not be within the scope of this work since our main focus is on the English language.

3. AI Technologies

This section explains the details of the AI technologies used with the approach from Section 2 to address the problem defined in Section 1.

3.1 Approach to AI

Our approach to AI in general is to leverage existing technologies where ever possible. Given that our solution is focused on natural languages (e.g. English) we have opted not to use any known NLP techniques. Many NLP frameworks and techniques are built / designed on the premise of understanding sentence structures ahead of time. In our approach, we have removed this from the equation allowing for a simple and pure approach with the intent that the system learns this over time. We do expect this approach to take more time to achieve.

It is important to note that there is a given recipe for this approach that permits for this capability. The details of the recipe are proprietary. However, the ingredients include the following concepts:

- Artificial Neural Networks
- Genetic Algorithms
- Knowledge bases
- Hybrid data stores (e.g. Graph, Document, Key/Value stores)

These ingredients are designed into a software architecture which is proprietary. Much of this architecture has been modeled after the known ways the human mind learns.

In addition to these ingredients, we will include our own unique forms of algorithms. These algorithms categorizations are below and are self-describing.

- Mutational algorithms
- Evolutionary algorithms
- Fractal algorithms

Lastly, we have a highly specialized form of number randomization which guarantees randomness absolutely. Random number generation is the heart of many of these algorithms and we have chosen to employ our own algorithms for this rather than leveraging existing techniques.

3.2 Technology functionality

The resultant functionality of our approach defined in section 3.1 is a truly 'free thinking" AGI which is undirected programmatically. It is directed through the training provided and any derivative learning which takes place as a result of this training. No two instances of the AGI are alike. This is guaranteed by the system architecture and the manor by which the algorithms are pieced together.

This AGI is capable of learning any language and scale to include images, audio, video over time. Scaling to other mediums include input and output of these media types.

3.3 Technology employed

We have discussed our proprietary approach with sufficient information providing a view into what will be developed. In section 2, we define the hardware and software underlying this solution.

We intend to evaluate cloud providers for multiple portions of the AGI. It has yet to be decided if these will be leveraged in the AGI. Moreover, we intend to explore a proprietary data center of our own as well as existing data centers (e.g. Oak Ridge, or standard data centers)

There are alternative technologies we may leverage listed below.

- MySQL
- SQLite
- Rabbit MQ 3.3.1
- nGinx
- Python 3.4.2
- Erlang R16B03
- Requests 2.3.0 (python library)
- Memcached Client 1.5.1
- Mysql Connector 1.1.6
- RabbitPy 0.19.0
- Celery 3.1.1
- Memcached Server 1.4.19
- Django (latest)
- Arango DB
- Open CL
- Zero MQ
- Numpy
- Scipy
- Scikit-learn
- Numba
- Sympy
- Pandas
- Matplotlib
- Theano
- Theanets
- Lasagna
- Keras
- Mozi
- Tensor flow
- Blaz
- Dask
- Datashape
- Mxnut
- Torch
- ROCm
- Caffe
- Arrayfire
- Mango
- Open AI – gym
- Anaconda
- Mahout
- Kafka
- Samza
- Apex
- Heron
- Storm
- Spark
- pyvalid
- jsonobject
- setproctitle

Some of these technologies are supporting functions while others are related directly to machine learning. Not all will be leveraged. There may be other solutions used not listed here. Lastly, we are flexible in the event that new technologies and solutions become available.

3.4 Validation and testing

Various forms of testing and validation will take place continuously throughout the life of the plan.

3.4.1 Unit Testing

With any software solution developed, unit tests are typically generated and or created to ensure proper functionality of all parts of the system. Our goal is to unit test as much as possible to ensure all underlying code is in proper working order. Unit tests will be a part of the continuous integration process.

3.4.2 Regression Testing

Similar to 3.4.1, regression testing will automatically take place during each build cycle ensuring previous code units work as expected.

Regression tests will be a part of the continuous integration process.

3.4.3 Performance Testing

There are key aspects where performance bottlenecks can be a concern. There are two specific areas we will focus on when it comes to this type of testing.

1. **Machine learning performance**

   To ensure optimal performance, we will base line the difference between AMD and NVidia based GPUs to ensure optimal performance. The only caveat to this would be the use of an agnostic software layer which may hinder performance. For example, Open CL may perform slower on NVidia GPUs when compared to using the native CUDA libraries.

   Any algorithm used (custom or not) will be optimized for peak performance allowing for execution across 1 or more GPUs including the spanning of servers of multiple GPUs.

2. **Graph and data store performance**

   There is a concern when it comes to using multi-modal data stores. Some data stores do not have the decades of optimization as do standard RDBMSs. We will start by choosing data stores that perform optimally in single server deployments. This will extend to include performance testing over clusters of

servers. Preference will be given to data stores built with native programming languages such as C or C++. Preference will also be given to those that have the most support, documentation and graph traversal languages.

Lastly, strict attention will be paid to graph traversals ensuring only the fastest of solutions are employed while optimizing all graph traversals for the absolute fastest executions.

3.4.4 Validation

The validation of the system proves to be the trickiest portion. Considering the unknown behaviors expected from our approach. The system may react differently between exact or similar input. The key to our approach of validation is to show signs of improvement and machine thought. Whether incorrect or not. Much like that of humans as we learn we make mistakes. Even after we learn we make mistakes or conclude the incorrect response.

3.4.4.1 Base cases

At the very basic of levels, the system receives input and determines whether or not to respond. Input and output is given reinforcement (e.g. positive, negative or neutral). Based on the input and reinforcement, the system determines what must be done.

There are a few basic workflows which will be tested. These workflows are modeled after the human minds way of learning. There are at least 10 different workflows which will be tested. Some of these are derivatives of the other while others are completely different.

In all cases, each work flow will have access to a few basic functions. These include storing information, fetching information, learning, training, and associations on data. The workflow dictates if and when these units of work are executed for the given context.

3.4.4.2 Evidence of learning

The evidence of learning can be validated by analyzing the workflows executing and whether or not the system chose to learn based on its inputs and outputs.

External validation will come in the form of a response. For example, repeating the word “mama” to a baby until they repeat the same word. The child may choose to respond or not. Eventually they conceptualize and associate the sound of “mama” with their mother. A similar approach is used with images

except that we are contextualizing the images as they arrive where required until the system is capable of understanding the image context on its own.

3.4.4.3 Evidence of thought

The system grows over time through these interactions stated in section 3.4.4.2. The system is designed to continually analyze the data it has absorbed and learned wherever it deems necessary.

Validation of thought works much like the validation of learning. We simply analyze if the system is indeed "rethinking" the data it has absorbed. Ultimately, we predict the system will begin to provide output as it sees fit regardless if there is interaction coming from an external source.

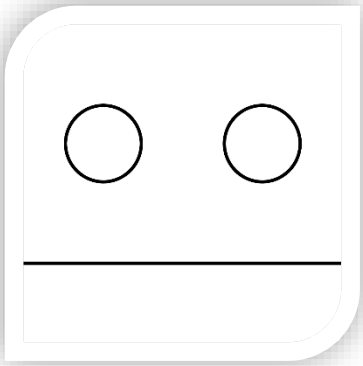
3.4.4.4 Evidence of association

Throughout the course of the life of the AGI we will determine if it is capable of associating words to words and concepts. A simple graph traversal and / or analysis of the graph will indicate if this is occurring. Moreover, association is a core function of the system. Therefore, it is safe to assume association provided any of the selected workflows execute and perform an action with its association component.

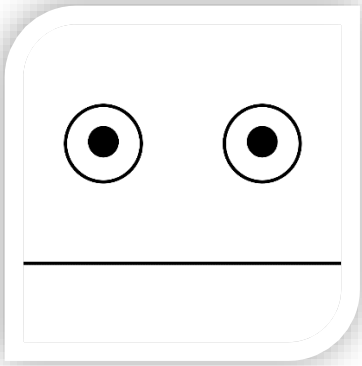
External validation of this will come in the form of a response provided by the system. Much like the example from section 3.4.4.2, eventually the system begins to associate additional words and concepts. It may choose to respond to a given input or without based on its association.

There are a few additional things we will use during the development, testing and validation of the AGI. We will add simplified user interfaces to provide quick and insightful statuses of the AGI during its execution.

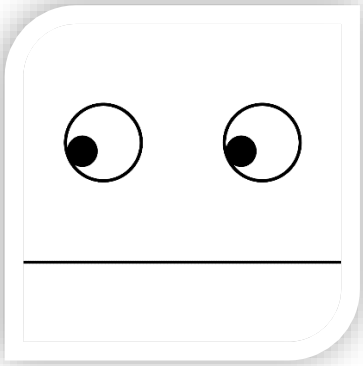
The following illustrations demonstrates the status of the AGI at any given time. We will use a very basic cartoon face which is comprised of only eyes and a mouth. The mouth will be used to indicate there is a response coming from the AGI and will move but not specific to the words produced. We may add eye brows to indicate the reinforcement type (e.g. positive, negative, and a neutral).



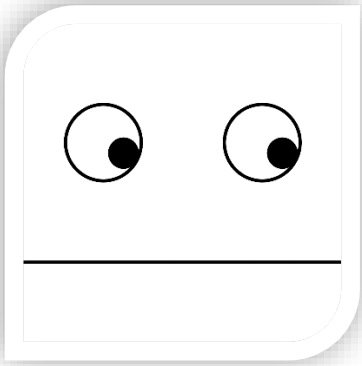
Inactive / Null state



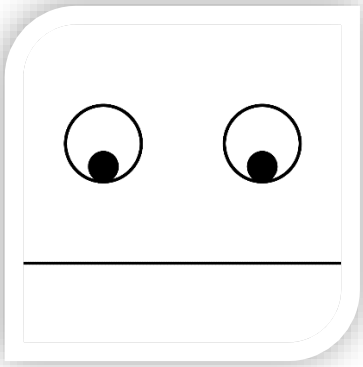
Stimulus or Reaction



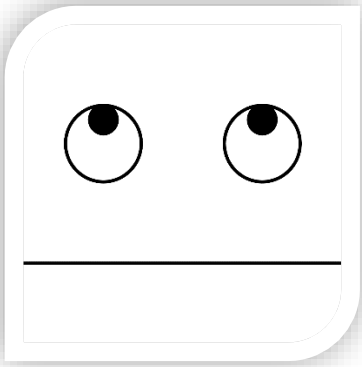
Learn



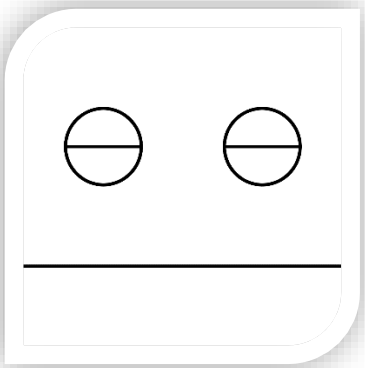
Associate



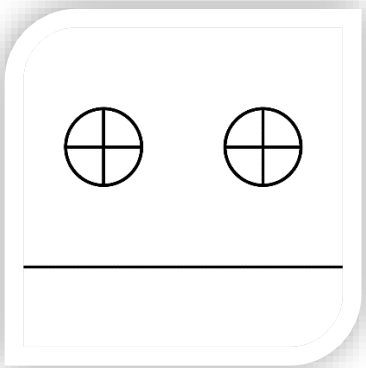
Remember



Invent / Create



Dream



Hallucinate / Day Dream

3.5 Related problems

There are many problems to be solved with an AGI of this kind. We will address the immediately obvious within this section.

3.5.1 Search engines

Modern day search engines are comprised of crawlers and indexers backed by proprietary page ranking and sorting algorithms. The problem with search engines is we treat them as if they were a person with all the answers. There are times when a search engine takes you straight to the answer. However, when you actually dig deep a search engine requires digging through countless documents and websites to find the information sought.

How efficient is this really? Not much considering that we can do much better. Backing a search engine with an AGI would produce singular answers to our questions rather than a slew of potential answers. It would enable the backend systems to these new search engines with the capability to truly learn and adapt over time. Ultimately, providing end users exactly what they were looking for.

3.5.2 Real world contextualization

One unknown frontier in the technology space is the true contextualization of the real world. Contextualization is simply the ability to observe the world and derive information from it. For example, Facebook probably has the world's largest facial recognition database. They have contextualized pictures with respect to people.

Many solutions attempt to provide actionable data from real world data. However, we have only begun to scratch the surface. These solutions are starting to use

machine learning that is highly specialized to just one thing and only one thing. There is very little room to do additional functionality.

For example, if raw video feeds from every airport in the world were coming into a central hub would that hub be able to detect if there is any sort of illegal activity going on? How would the system be able to determine the difference between a domestic violence situation and a couple arguing because they are late for a flight?

3.5.3 Automated assistants

A vast majority of the population has a smart phone (e.g. IOS or Android). These smart phones come equipped with assistants (e.g. Siri or Google Assistant). The problem with these assistants is that they quickly become useless. It really doesn't learn anything about us. At least not enough to be useful. Countless times people turn it off or get annoyed with this feature. This has been the case since Microsoft Bob or Clippie from MS Office.

What if there was an assistant that could actually provide real value?

3.5.4 Inefficient workforces

With all the modern advances in technology, little is done to improve manually performed work. That is, work that requires human intervention.

For example, construction sites have a wealth of systems at their disposal. They have design software that aids in building high rises, nuclear power plants and road ways. However, these jobs are still highly labor intensive. There are build time decisions that get over looked. What happens when these workers follow all the plans yet something still goes wrong? Thousands if not millions of dollars can be lost. Not to mention the time wasted on a project.

If an AGI was coupled with every worker on the site it will warn them during build time of issues that might arise. To be specific, laying down a foundation as per the plan when the plan was wrong. It may not have enough easement. Yet another example would be if power line polls have enough spacing between themselves and the building adjacent to it. Often times, these sort of issues are captured after the work has been performed.

This also applies across sectors and industries. It could be used for sales, business development, marketing, customer service, and technical support in much the same way. Possibly just a simple voice over on a phone call would suffice.

3.5.5 Learning aids

Our world has added a wealth of information at the click of a mouse button. Wikipedia, Google, FAQs and similar functionality have increased the speed of knowledge adoption on all forefronts. However, there are still times where these fail us.

A learning aid AGI would assist anyone in the learning process. Having coupled the AGI early on in the process with those attempting to learn a new topic the AGI would be able to watch the patterns of the users as they learn. Where they spend the most time and energy. More importantly, where they get lost in the information.

An AGI would be able to assist them by providing the answers needed at the right time. This is accomplished by simply introducing the information at the right time or prompting the user with a few questions. This would occur on its own.

3.5.6 News consolidation

Today we don't have any issues with a lack of information. We really have an issue with too much information. More importantly, too much invalid information. Look at the recent presidential election in the USA. There was a wealth of news from every facet. Some fake, some real – who really knows? There is the left side and the right side. Fake news, tabloids, real news all bunched into one.

How do readers know the difference? Users are left with options like Flipboard or other news consolidators. Or even worse, their Facebook feed. The problem with these solutions is that we are inundated with data all the time. The data repeats itself and it's literally thrown at you in any order. Most readers may opt to not even read the news at all. This is a common answer when asked if they read the news.

The AGI based solution for news aggregation would consolidate these news feeds and ads. Moreover, it could be used to rewrite the articles in an agnostic format. Citing different publications with different angles.

3.6 Training

Our approach to training is unique. We believe in training once and training at a hyper active rate. Moreover, we believe in letting the system retrain itself based on proprietary algorithms which identify areas of retraining. Lastly, we believe in letting the system articulate itself and attempt to generalize its learned trainings to enable faster training times while reducing the need for constant manual training.

3.6.1 Meta data

3.6.1.1 Initial data

Initial data used for training of the AGI will be entered manually. This allows us to validate the concepts and underlying components of the system. Unlike other systems, our AGI will consume bytes of data with related reinforcements to that data. Optionally, a desired reaction will also be provided to the AGI. Overtime, we theorize the AGI will learn based on this information and learn the ability to react, respond or stimulate external users of the AGI with this learned data.

This scales to include a crowd sourced model. The crowd sourced effort will leverage existing services like Mechanical Turk or related services. Users will be instructed to enter stimulus with reinforcement and desired reactions. The observations made here will validate the system is adapting and learning based on the stimulus it receives.

3.6.1.2 Bulk core data

In an effort to rapidly scale and train the AGI we will bulk load metadata used for training directly into the AGI data store. The system will later pick up and start its own retraining processes to start categorizing the information accordingly.

This metadata will be comprised of every know letter, word, and definitions of these items in the English language. Once loaded, the system will not be able to use this information until it has retrained itself on the data. This retraining effort occurs over time as stimulus / reactions tuples are fed into the system. The crowd sourced effort will continue during this part of the process to expedite the training.

3.6.1.3 Bulk customized data

The goal of this approach is to expedite the training more so than the previous approach in section 3.6.1.2. In this approach, we will analyze how the AGI categorizes the stimulus / reactions tuples. From this, we will create customized data sets used for training that will not require retraining or training. In a sense,

it would automatically impact the AGI at a grand scale and would immediately be able to make use of it in future stimulus / reaction interactions.

3.6.1.4 Bulk conceptual data

There will reach a point in the maturity of the AGI where it will become evident that all approaches have been successful and indeed it is accomplishing its goals. Our approach with this concept is to feed the AGI conceptual data by means of publicly available transcripts. These transcripts may be from conversations between individuals, movie scripts with dialog, or fictional literal works of art.

4. Evaluation Goals and Test Suites

This section describes how we should be evaluated in terms of assessing our progress in relation to the milestones achieved.

4.1 Evaluation and assessment guide

4.1.1 Basic assessment

This includes basic items related to the underlying mechanisms of the AGI.

4.1.1.1 System architecture

Architecture diagrams will be demonstrated illustrating the high level architecture of the system proving systemically the validity of the system.

4.1.1.2 System infrastructure and scaling

Hardware and infrastructure diagrams will be demonstrated illustrating the hardware and infrastructure required for the AGI. Moreover, it will focus on its ability to infinitely scale. Lastly, it will include a hardware to software mapping to fully understand the demands of the system.

4.1.2 Advanced assessment

This assessment includes the advanced features related specifically to the AGI and the validation of the system from the outside in. Judges will be provided a user interface which is conducive to these assessments. This may be altered to adjust to new found information either with the AGI or changes in recent technologies.

4.1.2.1 Capability to learn reactions based on stimulus reaction tuples

System will demonstrate a means to learn new or existing reactions based on any stimulus received into the system.

4.1.2.2 Capability to produce stimulus from its own stimulus reaction tuples

System will demonstrate its own reactions and stimulus produced of its own accord. The user interface will illustrate this to judges.

4.1.2.3 Capability to understand basic words and reactions to words

System will demonstrate a basic understanding of words and reactions to these words by means of its user interfaces.

4.1.2.4 Indication of independent thought

This is an observed assessment. It may occur whenever the AGI choses either during stimulus from judges or whenever it chooses. Furthermore, the AGI may also opt to not react to stimulus. This is also of its own free will.

4.1.2.5 Ability to produce more word structures

Continued interactions with the system will ultimately dictate if the system can indeed form more complicated responses or stimulus. This is to be observed throughout the evaluation of the system.

4.1.2.6 Ability to formulate thought

As with 4.1.2.5, this feature is an observation that will occur while the judges interact with the AGI through the provided interfaces.

4.1.2.7 Indication to adjust and improve over time based on any data available to the AGI

Another observed behavior of the system. This can be assessed and evaluated by producing similar stimulus repeatedly expecting either similar reactions or new reactions. This assessment may take time as the system would have to rethink, relearn, re-assess itself in order to adapt. It's unknown how quickly this will occur since it's determined by the system.

4.1.3 Post bulk data and bulk training assessment and evaluation

For this process, essentially repeat section 4.1.2. The difference is rather than using the AGI built upon the data entered manually through normal interfaces and the crowd sourcing efforts, the AGI has been bulk loaded in the form of raw data and training through rapid stimulus / reaction tuple insertions. Moreover, the AGI will process the new found information and begin to relearn and think through all associated data making connections between the data wherever possible.

This part is only possible if we can successfully achieve the goal of rapidly teaching the AGI instead of manually.

4.1.4 Social assessments

This section is related to the social media / outlets portions of the AGI.

4.1.4.1 Capability of receiving stimulus from social media interactions

Stimulus will be received through social media outlets. The user interfaces provided will indicate the receiving of stimulus. System monitoring will also provide proof of stimulus.

4.1.4.2 Capability of producing a reaction from social media interactions

Upon receiving stimulus from social media, the system will determine whether to produce a reaction or not based on stimulus. This can be observed publicly on the selected social media outlets.

4.1.4.3 Capability of producing stimulus / reactions autonomously

The system will also create its own unprompted stimulus to social media. The social media outlet, and those connected to it and the AGI may opt to respond or not. Judges may choose to interact with the system by means of social media during this assessment.

4.2 System of testing and assessment

As mentioned in section 4.1, judges will be provided interfaces to the system or documentation allowing for proper evaluation and assessment of the system.

5. Annual Progress Plan

This section describes and defines our progress plan. Our progress plan indicates our expected annual progress towards achieving our goals related to the problem stated in Section 1.

5.1 AGI Progress Plan

Year 1

CREATE THE AGI

1. Architecture and design documents created
2. Infrastructure build out (e.g. cloud servers, physical servers)
3. Development environments created (e.g. servers, workstations)
4. Object model draft 1
5. Data model draft 1
6. Sample data creation
7. Message definitions
8. Component development
9. Integration development
10. Data store development
11. **AGI Release Candidate 0**
12. Revised Object model draft 2
13. Revised Data model draft 2
14. Integration testing and validation
15. **AGI Release Candidate 1**

Year 2

EVOLVING THE AGI

1. Development of stimulus / reaction scripts
2. Crowd sourced engagement – scripted
3. Analysis and validation of approach
4. Improve training speeds
5. **AGI Release Candidate 2**
6. Crowd sourced engagement – unscripted
7. Analysis and validation of approach
8. Validate improved training speeds
9. Bulk load training seed data
10. **AGI Release Candidate 3**
11. Free flowing unscripted release of AGI to social environments
12. Continued analysis and validation of AGI

Year 3

AGI APPLICATIONS

1. Identify initial candidate use cases of existing work life
2. Identify initial candidate use cases of existing personal life
3. Design sample applications for demonstrations

4 Development sample applications for demonstrations
5 AGI Release Candidate 4
6 Release in isolated environments interacting with human counter parts
7 Analysis of effectiveness of AGI and human collaboration
8 Validation of effectiveness of AGI and human collaboration
9 Adjust AGI as required based on findings
10 AGI Release Candidate 5

For each year, it is expected that the plan may be changed slightly. We do not foresee year 1 changing dramatically. However, year 2 and 3 could be altered based on immediate findings.

We foresee additional years of ongoing development efforts with the AGI. As with any software system, it must be maintained, expanded, improved and changed to fit the demands of the future. The previously illustrated roadmap simply guides the reader and sets expectations for what is to come.

6. Grand Impact Goal

In this section we articulate the goal and the impact our AI solution will have on this goal.

6.1 Goals

The following section discusses and illustrates the overall goals for the proposed AI (AGI).

6.1.1 AGI Creation

This goal is to create an actual AGI that functions in a similar manor to how humans learn. The learning is focused on imagine contextualization, image identification, image training, and natural language without the need for known NLP techniques. The intent is to develop the AGI with the ability to learn and interact with images and the English language while leaving it extensible such that the addition of audio, video, and any other medium of data can be affixed to this.

For example, when the AGI learns the word "apple" – over time – it would relearn that the word "apple" appending to its knowledge base the image of an "apple".

6.1.2 AGI Validation

In this goal, we validate that the AGI is able to learn as humans do as well as show promise for scaling to larger contexts. For example, the ability to form parts or whole sentences. Little attention will be paid to grammar. The main focus is on the ability to group together words into a meaning of sorts.

6.1.3 AGI Human Collaboration

In this goal, we focus on the interactivity pieces of the AGI. Key questions to answer here are:

- Is it able to interact?
- Can it create a dialog?
- Will it be able to hold a conversation?
- How quickly will it respond to new found stimuli?

6.1.4 AGI Effectiveness

The effectiveness of the AGI addresses the viability of the AGI for all applications it may be applied to in the future. This will determine how effective it may be once incorporated human counterparts in the real world.

6.2 Importance to humanity

There are many areas in the world where the AGI will be impactful and important to humanity. Our goal within this section is to address the importance we foresee in the near and distant future for humanity. Take into consideration that this AGI evolves and grows overtime. Therefore, in the future it would be far more advanced due to its underlying architecture as well as advancements that may be made with hardware that would further improve the AGIs ability.

The following table breaks down a few of the areas we foresee being important to humanity:

Unemployment

Without proper care of this technology integrated into our daily lives we could see drastic impacts to the workforce. Specifically, the focus of workforce automation could render work forces of the past useless.

Understanding the technology from all aspects would allow us to better understand what the technology can actually do while determining how much of the technology is science fiction.

Inequality

The importance here is to understand how much inequality will be created by the creation of this AGI. Inequality in the sense that workforces would be automated creating gaps between business owners and their employees. The ripple effects of this could leave millions in despair if not properly handled.

Although this falls outside of what we can do in the near term - the topics of wealth creation and destruction are brought to light with this technology more so than technologies of the past.

Behavior

Humanity has always been impacted in some shape or form when it comes to any new technology being introduced into the equation.

How will humanity behave once an AGI becomes main stream?

Understanding how the AGI impacts behavior will be analyzed throughout as humans collaborate and interact with the AGI. This understanding provides a

Security

There are many aspects here when it comes to security in general. Is it related to cyber security? Is it related to physical security?

We believe the security lens applied to be slightly different. The lens applied here is ensuring the AGI is persistent and accurate to a degree where it can be used daily.

For example, an AI can be easily be fooled when it comes to computer vision. A few

glimpse into what is to come in our near future once AGI is in the hands of every person on the planet.

Learning

The learning process for each person in the world is both unique while similar at the same time. Some people learn one way while others learn a completely different way. The rate at which we learn is different between individuals as well.

The AGI coupled with a human through the learning process would be capable of understanding their human counterpart throughout the process while determining where the human is struggling most.

This would ultimately enable classrooms around the world whether virtual, home or school based with the ability to effectively assist the learning process for individuals regardless of topic.

Cyber Bullying / Harassment / Violence

These three topics are a part of human nature which is quite unfortunate. However, we have seen that the mass adoption of the internet has brought these topics onto a new stage. Now anyone can be effected across the world. For example, a sexual predator could be 10,000 miles away from a young girl. The internet connects them and could put the young girl in harm's way.

Current techniques have proved to be rather limited. What really defines this act? Simple key word searches could not be applied to this.

dots on a piece of paper could throw off results.

Our intent, is to ensure the AGI is capable of overcoming these nuances in the world.

Singularity

A fairly heated and much talked about topic in our world today. There is a lot of talk and speculation with far less analysis and study given to the reality of the situation.

What is an AGI? When does it become an ASI? Can it become an ASI? What happens after?

Throughout our efforts, we hope to gain a wealth of insight into a AGIs capabilities currently as well as the future of this technology. A proper AGI that evolves to an ASI which is understood would aid humans at the apex of this event should it ever occur.

Terrorism

Refer to Cyber Bullying section. The same applies here.

AGI would be very impactful here in regards to assisting with the detection of potential terror attacks and / or groups. The capability of understanding human languages in an intuitive manor implies that that AGI would be capable of determining intent regardless of specific word usage.

This would enable protection agencies with the ability to distill the vast amounts of communications taken place on the

An AGI understanding that there is a negative threat within a conversation with the understanding of the actors in place could serve as a red flag indicator to authorities.

This also lends to the "Turning Noise to Signals" coming up next.

Turning Noise to Signals

The internet is flooded with information. Some good, some bad, same fake, some real. The point here is that it's nearly impossible for us to deal with this abundance of information. How do we filter out all of this and distill it down to what we need or want to see?

An AGI would be capable of consuming mass quantities of information and distilling it down to actionable results. This would aid researchers tremendously. The countless hours spent mulling through piles of information would be reduced to the relevant information required for the researcher to perform their jobs.

internet to actionable results which could be investigated individually on an as need basis.

Our largest concern here turns out to be a double edge sword when it comes to matters of privacy and our rights as citizens. We do not intend to address this specific portion at this time.

Understanding the Public

There are various topics here related to each other. Much like the previous 3 examples (Turning Noise To Signals, Terrorism, Cyber Bullying), understanding the public is an important factor in our world. For government relations to business and families.

AGI is capable of providing this information in much the same way it Turns Noise into Signals by producing actionable results over mounds of data.

Space Exploration and Beyond

The dream of many science fiction aficionados. Exploring the unknown finding new life on distant planets is a part of this dream. There are many obstacles, problems and issues along our path to successfully exploring space let alone reaching alien life forms if indeed we find them.

We believe a technology that would greatly aid this effort is the use of AI. AGI would allow us to make space crafts without the need to support human life. To date, many of the probes sent out into space have been more of a drone than anything else.

Equipping the future probes with AGI would enable them to autonomously adapt to situations and explore based on specified scripts, frameworks and parameters.

Lastly, in the event that we do indeed find higher level life forms alien in origin – how in the world do we plan to communicate with them in their own language? An AGI such as the one proposed, equipped with the capability to learn spoken and written languages would be better at adopting the new languages. These AGI would act as a proxy, or rather, a universal translator of sorts in our efforts with our communications.

7. AI Human Collaboration

In this section we will describe how our AI solution will work with humans to achieve the desired results.

Throughout the entire Plan we have made mentions to how our effort is combined with human collaboration. The following table readdresses these topics commented on previously.

Base Collaboration

- Initial testing and seeding of base test cases from engineers
- This process validates the stimulus reaction capabilities of the system
- Human interaction is performed here to validate the success of learning over time

Indirect Collaboration

- Comprised of compiled data from humans
- Used to seed initial stimulus / reaction tuples
- Derived from human interactions
- Based on the known understanding of how humans learn as infants

Crowd Sourced Collaboration

- Scripted in nature initially
- Crowd sourced workers (e.g. Mechanical Turk) is leveraged for this process
- Once manual stimulus reaction scripts are performed by humans the humans are instructed to interact naturally
- Objective is to direct each crowd sourced resource is an objective and a proposed script or script framework to follow for interaction with the AGI

Social Collaboration

- Unscripted interactions
- Leverage social media outlets
    - Facebook
    - Twitter
    - Email
    - Slack
    - SMS
    - Reddit
- Only possible after previous collaborations take place
- AGI continues to expand and grow with these stimulus / reaction tuples