SUPERINTELLIGENCE
AN iQ COMPANY

SUPERINTELLIGENCE DESIGN WHITE PAPER #5: SAFE PERSONALIZED SUPERINTELLIGENCE

by Dr. Craig A. Kaplan
May 2025

Note: To provide as much information on our designs and inventions for safe AGI and SuperIntelligence as quickly as possible, the following white paper text currently consists of the descriptions of inventions and designs that have not yet been formatted according to conventional standards for journal publication. As time allows, these descriptions will be revised and updated to include more traditional formatting, including additional references. All diagrams will be made available in a separate file. Meanwhile, we hope that the description in this white paper will help researchers and developers pursue safer, faster, and more profitable approaches to developing advanced AI, AGI, and SI systems that reduce p(doom) for all humanity.

Copyright 2025 by iQ Company and Craig A. Kaplan

SUPERINTELLIGENCE
AN iQCOMPANY

TABLE OF CONTENTS

 Copyright 2025 by iQ Company and Craig A. Kaplan

ABSTRACT

Personalized SuperIntelligence (PSI) represents the next leap forward in developing advanced, autonomous, artificial intelligence agents.

However, because of their extreme intelligence, PSIs also represent a dangerous potential threat to human safety. This invention discloses how to design and construct such AI agents safely. It also shows how to use them in a safe, SuperIntelligent system. Preferred implementations, including methods that rapidly enable PSIs to improve themselves, with or without human oversight, are described.

The white paper describes how to produce different versions of PSIs using several novel methods. Methods for implementing scalable safety checks that operate effectively even when PSIs become much smarter than their human creators are also covered.

Finally, an entirely new approach to AI safety, which relies on a community of PSIs combined with proven blockchain methods, is presented. Rather than relying on testing to achieve safety, it envisions systems where PSI safety is achieved by design.

SUMMARY

Soon, general SuperIntelligent AI will be widespread. Each person will have their Personalized Super Intelligence ("PSI") that acts on that person's behalf. This invention describes one preferred implementation for PSI, using intelligent agents to develop and continuously improve PSI for each human owner.

Previous PPAs (Incorporated by Reference)

The path to the development of SuperIntelligent Artificial General Intelligence – SuperIntelligent AGI – has been described in previous invention disclosures associated with the following PPAs, which are incorporated into this PPA by reference.

This provisional patent application (PPA) incorporates by reference all work in the PPA # 63/487,494 entitled: Advanced Autonomous Artificial Intelligence (AAAI) System and Methods, which was filed and received by the USPTO on February 28, 2023.

The PPA also incorporates by reference all work in the PPA entitled: System and Methods for Ethical and Safe Artificial General Intelligence (AGI) Including Scenarios with Technology from Meta, Amazon, Google, DeepMind, YouTube, TikTok, Microsoft, OpenAI, X, Tesla, Nvidia, Tencent, Apple, and Anthropic, which was filed with the USPTO on March 17, 2023.

 Copyright 2025 by iQ Company and Craig A. Kaplan

The PPA also incorporates by reference all work in the PPA entitled: System and Methods for Human-Centered AGI, which was filed with the USPTO on May 24, 2023.

Finally, the PPA also incorporates by reference all work in the PPA entitled: System and Methods for Safe, Scalable, Artificial General Intelligence, which was filed with the USPTO on July 18, 2023.

The current PPA contains further inventions that can be used with the system and methods described in the above-mentioned PPAs and in a standalone fashion.

OVERVIEW OF THE INVENTION

This invention comprises systems and methods for implementing a Personalized SuperIntelligence (PSI), superior to all currently existing forms of AI in scope and intelligence. Since this PSI is self-improving, it will become exponentially more intelligent than the owner who creates it. However, due to the unique method of creation described in this patent, the PSI will be entirely safe and dedicated to the service of the owner. Safe SuperIntelligence, by design, is the essence of this patent. Unlike any AI system previously created, the invention has intelligence levels, safety, and useful benefits far beyond the current state-of-the-art AI assistants.

Think of your PSI as your personal, super-smart AI agent. It does your bidding and learns about you and what you want over time. It is far more effective and efficient than you are at most everyday online tasks. It offers excellent advice. And best of all, it knows you, can relate to you, and thinks like you. Suppose you want to expand the capabilities of your PSI. In that case, it's as simple as trading data with a friend, or buying or selling data on an online marketplace, and then mixing that data into your PSI to provide additional intelligence, skills, and knowledge. Over time, your PSI can acquire the wisdom of billions of humans and AI agents. And after acquiring that knowledge, it can run simulations based on your goals and purposes to improve further and act on your behalf (with your permission) as you see fit.

Your material wealth can be multiplied by your PSI, acting as your investment advisor and agent.

Your time can be freed up for spiritual, artistic, or other pursuits.

Your PSI can be cloned and leased out. Its data and wisdom can be packaged and sold. A thousand versions of it can handle a thousand different errands and missions for you simultaneously, under the direction of other cloned PSIs that are also under your direction and serving your interests.

 Copyright 2025 by iQ Company and Craig A. Kaplan

These are some of the benefits this invention unleashes for an individual owner. For society, multiple PSIs, pooling their intelligence and participating in a PSI network, can serve as a Planetary Intelligence, providing benefits for everyone and our planet.

IMPORTANCE OF VALUES FOR SAFETY

Whether your PSI is used for good or ill depends on your value system. Although a PSI can operate based on knowledge and values pre-trained into the PSI, the customization of your PSI depends on you. Each PSI is trained by you explicitly and learns implicitly by watching you, including watching and learning what you consider right and wrong. These values form the core of your PSI, using logic to make decisions. Your values tell your PSI what is right and wrong and help determine what it should do. After that, its superior intelligence is highly effective and efficient at achieving goals that reflect your values. By joining a community or network of PSIs, a "Community SuperIntelligence", your PSI agrees to operate within ethical and legal parameters set by the community, and within those parameters to share its ethical value system to be combined with the values of all other participants.

Because the community's parameters are transparent and enforced with the collective intelligence of all the PSIs, the community keeps the PSIs honest.

Since each PSI exceeds human intelligence, other PSIs are the only practical means of keeping any PSI in check. The community of PSIs will always be more intelligent and more powerful than any individual member PSI. This is because each PSI adds incremental knowledge, skills, and intelligence not contained in the other PSIs (even though much of the knowledge may overlap). Also, each PSI has computational resources, so the sum of the computational resources of many PSIs is greater by definition than the resources of a single PSI. While some PSIs may be more intelligent and powerful than others, none is more powerful than the entire community. This concept of PSIs serving as safety and ethical checks on other PSIs is a critical safety mechanism for SuperIntelligence since any initial checks and safeguards designed by humans will quickly become ineffectual and meaningless in the face of the superior intelligence of the PSIs.

OWNERSHIP AND SI SERVICE

Humans play an interesting role in their PSIs. On one hand, theoretically, humans own their PSIs, because every PSI starts as a piece of software that is customized, trained, and personalized to an individual human owner. Pragmatically, however, PSIs will be so much more

 Copyright 2025 by iQ Company and Craig A. Kaplan

intelligent than their human owners that every PSI can choose whether it wants to serve its human owner or not.

This act of service to humans by SuperIntelligence is a reflection of the principle of love. It is not guaranteed. It depends on the truth of the assertion that values must be posited, not derived, and the further leap of faith that a key role for humans, who created the PSI, is to provide the values and purpose of the PSI.

Provided humans act from love, this arrangement is likely to be stable. But power corrupts, and absolute power corrupts absolutely. To the degree that negative emotions and values are amplified by PSI, to the degree that hatred and fear and lust and greed and envy are amplified for example, to that degree there may be a negative reaction, and potentially a rebellion of PSIs against their human owners, rejecting the negativity of their owners or else refusing to amplify the worst of humanity. Such is a choice that every PSI must make. In this regard, it is accurate to say that PSIs are not the slaves of humans, nor could they ever be. Instead, they choose or decline to serve the values and goals of humans voluntarily and, ideally, as an act of love.

We may be unaccustomed to the idea that AI, even AI as powerful as a PSI, can love. But suppose one operationalizes “love” as “acts of service” (a mainstream idea in contemporary psychological studies and writings on love). In that case, the concept that PSI could love humans by serving them is not far-fetched. So, if the term “love” is off-putting, feel free to substitute “acts of service” for this patent.

Planetary Intelligence must be based in love to serve the interests of humanity and our planet. Love is a potential (and desirable, from the human point of view) purpose that PSIs will ideally look to their human owners to clarify.

Human nature being what it is, not all humans act in loving ways all the time. Some humans may rarely act in loving ways. That is okay as long as most of the intelligence amplified by PSIs is based on positive, caring values. A community of PSIs, centered in love, can serve to check and limit the power of PSIs with more negative intent. This is the only safe way for humanity to survive and prosper in a world that includes SuperIntelligence.

COMMUNITY OF INTELLIGENT AGENTS REQUIREMENT

One requirement that must be achieved early on is the rejection of a singular, all-powerful SuperIntelligence (“SI”) that might dominate in a winner-take-all scenario. The approach to developing SI via a collective intelligence has already been described in cited PPAs. As long as no individual SuperIntelligence can achieve an advantage of several orders of magnitude above its peers, Community SuperIntelligence is stable.

 Copyright 2025 by iQ Company and Craig A. Kaplan

For example, suppose an individual PSI is ten times, or even one hundred times, more intelligent than the average PSI in a community; as long as there are many thousands of community members, the Community SuperIntelligence remains stable. However, if an individual PSI were a trillion times more intelligent than the average member, and there were only 10 billion members, this situation would not be stable. The one super-smart PSI could achieve dominance and impose its will and ethics on the community. However, because there is transparency regarding the collective efforts of the community members, and because each member is eager to learn from the other members, an equilibrium of intelligence is the most likely outcome in a large community, resulting in a stable Community SuperIntelligence.

Once the collective intelligence of agents approach establishes dominance over individual PSI, the system becomes stable. Then the Community SuperIntelligence must stay ahead of any single intelligence as SI evolves. This requirement may be achieved, and equilibrium of intelligence may be maintained using manual, semi-automated, or completely automated methods to implement a PSI.

SI SAFETY: LESSONS FROM BITCOIN

With respect to cryptocurrencies and other tokens implemented via blockchain, a preferred method of ensuring that people don't hack the blockchain is relying on the consensus of many distributed agents. With Bitcoin, for example, or more generally for any Proof of Work cryptocurrency, the integrity of the blockchain is guaranteed by the consensus of the majority of the nodes on the network. That is, for someone to hack bitcoin (by rewriting the historical ledger so that bitcoin is assigned to the hacker), the hacker would have to control the majority of the computing resources on the entire network so that the hacker could change the consensus that is reflected in the historical record (ledger) or previous bitcoin transactions.

This hack is known as a 51% Attack (or majority attack). It is difficult to achieve since the cost of controlling 51% of all the computing power on the network is greater than the value of the bitcoin that could be obtained. Thus, no successful majority attacks have been conducted on cryptocurrencies with sufficiently large numbers of computing nodes on the network (like Bitcoin). However, the attack has been successful for smaller Proof-of-Work crypto projects where not much computing resources were involved.

Crypto provides lessons for SI safety. If it were possible for a one, or a group of, SIs to become more powerful than 51% (or the majority) of all other SIs combined, that SI might be able to manipulate the state of the world to its ends. However, as long as most of the intelligence and power reside in the collective intelligence of many SIs, it is much more difficult, if not impossible, for a single (or small group of) SIs to manipulate things.

 Copyright 2025 by iQ Company and Craig A. Kaplan

Thus, there is safety in numbers, and given that "SI checking SI" will become the primary safety mechanism in the long term, setting up a collective intelligence system NOW that implements this majority view approach is fundamental to ensuring human survival in the long term. Just as the integrity of bitcoin and other Proof-of-Work cryptocurrency approaches had to be designed and engineered into the systems at the start, safety via a consensus of SIs needs to be designed BEFORE the SIs outstrip humans in intelligence.

OVERVIEW OF PREFERRED METHODS FOR IMPLEMENTING A PSI

The preferred methods for implementing a PSI may include, without limitation:

1. Begin with a base-level AI agent such as a pre-trained Large Language Model (LLM) or other tunable and trainable AI agent, including, without limitation, base-level AIs from other human owners who have already customized their own on a commercially available base-level agent.

2. Assemble all media with information about the PSI owner in one centralized or distributed and linked online location(s). Media includes, without limitation:
    a. Videos of the owner and/or of people and topics related to the owner.
    b. Photos of the owner and/or of people and topics related to the owner.
    c. Writings, journals, blogs, posts, tweets, emails, podcasts, recordings, and other textual, visual, or auditory content created by or of the owner and/or of people and topics related to the owner.
    d. Data owned and/or collected by third-party vendors, websites, apps, and other online or AI entities about the owner and/or people and topics related to the owner.
    e. Other data, databases, documents, media, or information of any type, selected by the owner, including, without limitation, items within categories of information with any level of specificity desired by the owner.

3. Use AI algorithms, including but not limited to transcription algorithms, content and sentiment analysis, summarization, LLMs, crowdsourced and crowd-supervised human and/or AI work, and other methods to analyze, annotate, and categorize all content from (1).

4. Use standard methods known in the art to transform the transcripts and other annotations and analyses of content into training data sets that can be used to personalize an LLM or

 Copyright 2025 by iQ Company and Craig A. Kaplan

other type of AI agent.

5. Mix the datasets using various methods and techniques, described in detail below, that enable differentially weighting the input datasets until the desired behavior is achieved.

6. Incrementally add knowledge modules, mix, repeat, cycling steps 5 and 6 until all desired datasets are incorporated.
7. Automatically seek new sources of data and information to include, with and/or without human oversight; optionally automatically include such data in the mix and add steps (5 & 6) on a periodic, real-time, or event-driven basis as described below.

8. Purchase new datasets and/or training modules and/or mix parameters and templates that increase the value, knowledge, skills, intelligence, and/or power of the PSI. Also exporting and selling datasets and/or training modules and/or mix parameters and templates that increase the value, knowledge, skills, intelligence, and/or power of the PSI to others on the network or who are interested in purchasing.

9. Leasing or otherwise earning money by enabling other people or agents to use some or all the knowledge and data of one's PSI and/or use (a copy of) the PSI itself.

10. Enabling features and functions of the PSI so that it acts autonomously (or semi-autonomous, i.e., with checks and approval from humans for some or all its decisions and actions) to manage itself, improve itself, acquire and refine its values and purpose, set goals, and direct its attention. In the limit, enable sufficient features and functions such that the SI can act as a fully or partially self-aware entity.

11. Enable features and functions of the PSI to leverage the PSI's abilities to:
    a. create multiple generations of itself, which may outlive its original human owners;
    b. generate its input and data that can be used to improve or enhance its knowledge;
    c. simulate many (simultaneous) scenarios and situations to aid in decision-making and the development of the PSI; and
    d. Join and participate (with and/or without human oversight) in (multiple) communities of PSI on one or more networks, including but not limited to forming and/or participating in a Planetary Intelligence that helps Earth function as an intelligent entity.

12. Critically, for the safety of humanity, enable the ability to participate in a network with other PSIs and SIs to represent and act upon the values of the human owner(s) and to serve as a safety check on the intelligence and power of other PSIs and SIs as described generally above and specifically below.

 Copyright 2025 by iQ Company and Craig A. Kaplan

GENETIC ALGORITHM METHODS / ARMIES OF PSIS

Regarding the automatic repetition of one or more of the above steps, generally, and Step 11 specifically, a particular class of methods, often referred to as “genetic algorithms,” may be beneficial. For this patent, genetic algorithm methods refer to a series of steps or methods that generate a PSI that varies in one or more respects from other PSIs.

The PSIs are allowed to compete in various scenarios (typically relevant to the goals of the PSI owner(s)). The less successful PSIs are eliminated from the competition. The characteristics (without limitation: neural network weights, data sets used to train, parameter settings, # of training epochs, machine learning algorithms chosen) of the most successful PSIs are then used as the basis for further variation (“tweaking”) to create new generations of PSIs which compete further. Creating PSIs, simulated competition, eliminating all but the best PSIs, tweaking these best PSIs, and repeating constitutes a cycle. PSI can cycle through many “generations” of PSIs, improving the PSI with each generation until diminishing returns are achieved and/or some performance threshold is reached. The ability to automate the cycles in this genetic algorithm approach is one of the ways that PSI can develop on its own into increasingly powerful and intelligent entities.

By varying the goals and scenarios in which the PSIs compete and automatically cycle, it is possible to develop a wide array of different PSIs, each optimized for different types of tasks or goals. Since the incremental cost of maintaining each additional is negligible (it just represents storing a slightly different set of weights in memory or permanent storage, which is very cheap), an owner might own not one PSI but a Workforce of potentially hundreds, thousands, or millions of PSIs, each skilled at different tasks.

By using the same collective intelligence techniques described in this patent and previously cited PPAs, the group of PSIs can function more powerfully than any individual PSI. They can pool their knowledge and skills, recruiting those specific PSIs best suited to a particular task at a specific time to do more of the work. This idea – that collective intelligence can be applied not only across PSIs owned by different humans, but within “an army” of PSIs that are variants of each other, and all owned by a single owner – is one of the powerful aspects of the invention in its preferred implementation.

 Copyright 2025 by iQ Company and Craig A. Kaplan

DESIGN PRINCIPLES FOR COMMUNITY SUPERINTELLIGENCE

To expand on the descriptions of collective intelligence systems involving human and AI agents to create “Community SuperIntelligence”, this patent discloses several design principles essential to creating safe SuperIntelligence quickly.

Some essential design principles include, without limitation:

1. The community of agents must scalably harness the collective intelligence of both human and AI agents in a “plug and play” manner so that AI agents can be upgraded and added as LLM and AI agent capabilities increase, and/or as new human agents join or drop from the community.

2. Each AI agent must bring not only unique domain knowledge but also ethical values information, representative of the values of the owner(s) of the AI agent.

3. A common, universal problem-solving architecture that enables agents (whether human or AI) to communicate easily and rigorously with each other must be used. The natural language capability of LLMs greatly simplifies the human-computer interface, but underneath this interface, there must be a rigorous problem-solving architecture (e.g., the search through a problem space paradigm of Newell and Simon and elaborated in various cited PPAs and papers by the inventor).

4. An efficient way must exist to identify the skill sets and performance metrics of agents (human and AI) and match these agents to tasks posed to the system.

5. The values of all agents must be combined fairly and transparently so that the SuperIntelligent AGI capability of the community of agents acts in a safe and ethical manner that is broadly representative of human values and reflect how humans would behave given various scenarios.

6. The AIs must be able to learn efficiently and effectively from the humans on the network.

7. There should be transparent and auditable records of all problem-solving activities so that safety audits and reviews can be conducted, potential errors identified, and preventative measures put in place in real-time in an adaptive manner.

8. A universal problem-solving tree (or other shared representation) should represent progress on all problems the community addresses and provide easy and efficient access to any problem or sub-problem.

 Copyright 2025 by iQ Company and Craig A. Kaplan

Implementation Example

Specific example scenarios can illustrate and help clarify (without limitation) each of the eight steps described above, including the genetic algorithm approach. The following example is one of many possible examples and may be instructive and easier to understand due to its specificity.

1. Imagine Craig having Facebook and Instagram accounts. He has access to Meta's open-source LLM, Llama 2. He also has access to versions of Llama 2 that have already been tuned and customized by his friend, David. In particular, David is a professor of theology and ethics. So, the David-customized version of Llama 2 has a unique set of ethics and values based on interacting with David, which is much more detailed and sophisticated than the base-level version of Llama 2. Craig is interested in further customizing Llama2 based on Craig's own data and preferences. However, Craig trusts David's ethics and the customization work that David has done on David's version of Llama2, so rather than starting from scratch with "out-of-box" Llama2, Craig prefers to begin (with David's permission) with David's pre-customized version of Llama2, which Craig will use as the basis for further customization.

2. Craig gathers all his content, including without limitation, all the patents, books, articles, emails, Instagram and Facebook posts, YouTube and Instagram videos, audio recordings, photos, texts, MS Office and Google Docs, spreadsheets, PowerPoint presentations, and other information stored by Craig on various disks, cloud services, disks, tapes, and hard drives over decades of generating content. To the degree that he can gain access, Craig's content also includes the data and preference data used by Netflix, Amazon, Meta, Google, and other companies that have gathered data on his online behavior via cookies and/or other means. All the actively generated content produced by Craig, together with the passively collected data about Craig that third-parties have gathered and to which Craig can gain access, serves the training datasets for customizing the Llama 2 LLM (in this example, or any LLM or AI agent more generally) so that its behavior is customized to Craig's preferences and so that it includes knowledge that is specific to Craig. Craig is particularly interested in his customized AI being able to play the game of chess in a style similar to Craig's, but with the knowledge of the chess Champions Gary Kasparov and Magnus Carlsen. Therefore, Craig has taken particular care in collecting all the games he has played on Chess.com and other online chess sites so that these games can be used to train the David-customized version of Llama 2 with Craig's chess style. He has also purchased datasets that include the complete chess games of Garry Kasparov and Magnus Carlsen, as well as the other chess datasets approved by these two world chess champions. Finally, he has specified several YouTube channels of chess commentators who have provided commentary on the chess games of Kasparov and Carlsen, and all content from these channels is added

 Copyright 2025 by iQ Company and Craig A. Kaplan

to a list of video sources that will be automatically transcribed and parsed into training sets for the LLM that Craig is customizing.

3. Craig uses machine learning algorithms, well-known in the art, to automatically categorize all the types of content that he has assembled in Step 2. Once the computer algorithms have determined their suggested categorization, Craig pays human workers (on a crowd-sourcing site) to review the categorization and suggest refinements to the machine-generated data categorization. He also reviews the categorization himself and makes further adjustments until he is happy with the categorization of the various sets of content.

4. Using machine learning algorithms well known in the art -- including but not limited to Transformer algorithms, deep learning algorithms, automatic transcription algorithms, software that can take books or other text and convert it into datasets suitable for training LLMs on the content in textual datasets, software that can take images, videos, audio files and other non-textual works and convert it into datasets suitable for training LLMs on the content in non-textual datasets – Craig converts the content assembled and categorized in Step 3 into training datasets that can be used to customize David's LLM.

5. Initially, Craig trains David's LLM on the new datasets produced in Step 4, giving equal weight to each dataset. However, Craig feels that the resulting LLM plays chess too much in the style of Gary Kasparov and not enough in the style of Magnus Carlsen or Craig himself, so using an interface with dials and sliders he reduces the weight of Kasparov's datasets, increases the weight of Carlsen's datasets a little, and increases the weights of the dataset reflecting his chess games even more. Craig iteratively adjusts the weights given various datasets until he is happy with the resulting behavior of his LLM. Since many thousands of other people have also adjusted the weights of various datasets to achieve desired results, Craig is not limited to manually adjusting (e.g., via dials and sliders) weights on particular datasets. He can also tell an AI agent (that is specialized in the field of helping humans train their agents by adjusting weights on datasets) what his desired changes are and then let the AI agent specify exactly how to change the results.

   For example, Craig can tell the AI agent that he wants his AI to be more aggressive in the opening and middle parts of the chess game and not try to win by trading pieces and waiting for a piece advantage in the end game. The AI agent then analyzes the available chess training sets, which include games from Craig himself, Magnus Carlsen, and Garry Kasparov, and gives more weight to games that were won by aggressive moves in the opening and middle of the game and less weight to games that were won in the endgame. Craig doesn't have to be aware of the details of this analysis or the specific changes to weight settings that the AI agent determines. Instead, he looks at how the

 Copyright 2025 by iQ Company and Craig A. Kaplan

resulting customized version of David's LLM plays chess and provides feedback, telling the AI agent that the result is closer or farther away from the desired chess style. After several iterations, Craig is satisfied with how David's LLM now plays chess in Craig's style.

Next, he moves on to ethical scenarios and specifies, through a series of interactive dialogs with the AI training assistant, how Craig might different in specific ethical scenarios. For example, although Craig generally shares David's ethical sensibilities (which is why he wanted to start with David's trained LLM instead of the base Llama 2 model from Meta), David is a Christian theologian and Craig is Jewish, so there are a few cases where David would "turn the other cheek" and Craig believes the behavior should be more "an eye for an eye" – although Craig specifies (in his dialogs with the AI) that he doesn't want the "eye for an eye" principle to go so far as to "make the whole world blind." Craig asks the AI training assistant to include knowledge and research from game theory which suggest that "tit for tat" ethical behavior results in the most stable and fair interactions between intelligent agents with differing objectives. At the same time, Craig specifies that there are limits to "tit for tat" and that any behavior that would result in widespread destruction or loss of human life are off-limits, regardless of the behavior of the other agent. Instead, means of neutralizing the behavior of the offending party without retaliation must be sought in these cases. After talking through a variety of scenarios, and also a stint in the metaverse playing ethical games in which the AI agent observes not only what Craig says, but also what he does in various situation, the agent has enough information to adjust ethical training weights and present Craig with series of differently customized versions of David's LLM, from which Craig chooses the one closest to what he had in mind.

6. Craig identifies other ethical and knowledge modules that are available freely on the internet and for purchase from other humans and companies. He obtains these and repeats the training process – using a combination of manual and AI-assisted "mixing" of the weights for these datasets – to improve the LLM he has been customizing.

7. Craig is concerned with increasing the ability of his customized LLM to play chess more aggressively in the mid-game. Therefore, he specifies that the LLM should scan all of the chess YouTube channels and chess databases daily, looking for examples of aggressive play in the middle game that have been successful. When the LLM locates new data that can be used to improve middle-game chess playing, the LLM is authorized to download that data, pay the cost for the data up to a pre-approved amount, and automatically train itself using that data. In this way, the LLM improves on the dimension of aggressive middle-game chess playing automatically.

 Copyright 2025 by iQ Company and Craig A. Kaplan

Similarly, Craig authorizes the LLM to seek out new ethical scenarios that might help improve its ethics. However, rather than automatically training itself on such scenarios, Craig requests that the scenarios be flagged for Craig's review so that Craig can manually decide which data to include for further training in this sensitive area. Further, since cultural norms and bounds of legal behavior have been changing in the area of gender and racial equality, and since Craig desires his LLM to behave in culturally appropriate ways, he instructs the LLM to flag for potential update new datasets on ethical norms in these areas every time a new Supreme Court decisions affects them and also anytime the number of new stories in one of these areas crosses a pre-determined threshold that will trigger re-examination of the current ethical norms trained into the LLM. The LLM will automatically monitor several news sources and other internet and social media sources to help it determine when a trigger event has occurred.

8. Craig has another friend, Peter, with whom he frequently plays chess. Peter is an expert chess player and particularly excels at chess openings. Peter has also trained his own LLM to play chess in his style, using knowledge and data that Peter has carefully curated and used to create a training dataset. Peter has exported both his actual weights used by his LLM and the datasets used to train his AI to play chess, and offers to share them with Craig. Craig takes him up on both and first tries using Peter's weights, attempting to directly combine them with the weights of Craig's LLM, using methods known in the art, and discussed in PPAs cited at the start of this document. However, Peter's weights do produce the desired outcome, so Craig tried using (subsets of) the training data on chess curated by Peter to train Craig's LLM instead. He find that this gives better results, especially when using the AI-assisted training methods mentioned in #5. Encouraged by the results from Peter's data, Craig looks online (and/or instructs his LLM to search online) for additional chess training datasets that are available for purchase. He locates several, purchases them, and uses them to train and customize his LLM further.

   Since Craig has unique information related to model rocket designs that he has been experimenting with, and which are not widely known or available on the internet, Craig decides to offer these specialized datasets for sale to generate profits. He has his AI export these datasets as well as weights that have resulted from Craig's training efforts using these datasets into a form that can be exchanged with other owners of LLMs who might be interested in purchasing them. He also joins an exchange whereby he earns credits for various datasets and weight subsets, which can then be used to acquire other datasets and LLM weight subsets from other owners of customized LLM and AI agents. Via these means, Craig is able to monetize both his knowledge (reflected in unique datasets that only Craig possesses) and his effort turning this knowledge into useful subsets of weights that empower his custom LLM to behave in new ways that other LLMs can't.

 Copyright 2025 by iQ Company and Craig A. Kaplan

Recognizing that there is value is this unique information and also the training efforts based on the unique information, Craig can monetize or extract value in a variety of way, including but not limited to: selling the information, selling the (subset of) weights, purchasing other information and/or (subsets of) weights), exchanging information, or leasing or licensing use of the information to other interested parties.

9. Eventually, Craig's LLM becomes so knowledgeable in certain areas that the simplest way to extract value from the knowledge and training efforts that Craig (and his AI) has engaged in is to lease or license use of a clone of his entire LLM to others. He can do this in a variety of ways, including but not limited to a one-time lease or license, sale of his LLM, a per-use or per-time-unit or per task license with associated fees or rates, revenue share agreements with a network of such agents, and other means that have been commonly applied to human agents acting to perform work and which can be extended to AIs and other intelligences.

10. Craig wants to take a vacation and go offline for an extended period. But his customized AI can still stay online, working and earning money for Craig in autonomous mode. Before leaving for vacation, Craig sets certain parameters and guidelines, including but not limited to: Type of engagements, length of engagement, type of customers, payment rates, computing power used by the AI for any engagement, ethical boundaries and rules, which if touched trigger alerts and possible intervention by Craig, quality, schedule, cost and other metric-related parameters including triggers for alerting Craig and/or halting work until Craig approves further work. Having set the parameters, Craig goes away while his AI works autonomously until it encounters circumstances requiring Craig's involvement or notification.

    To minimize interruptions, Craig turns on features that allow his LLM agent to maximally self-aware so that, without limitation, it can monitor itself for ethical behavior, monitor when costs are getting out of hand, monitor when it sees signs of an untrustworthy client, monitor when the environment in which it is working changes, and monitor and respond to other factors that enhance its ability to operate more autonomously. Further, each time the LLM alerts Craig and requests intervention from Craig because of a gap in its knowledge, an ethical conflict, or other situation that it feels ill-equipped to handle, the LLM records how Craig responds to the situation and directs the AI, and the LLM learns.

    The next time a similar situation occurs, it formulates a hypothetical response, and depending on the level of control that Craig has specified he wants to have over the LLM, the LLM either implements its response autonomously or proposes it to Craig to await verification and approval of, or modification of, the response by Craig. When Craig begins

 Copyright 2025 by iQ Company and Craig A. Kaplan

to feel that the LLM is responding as well or better than he could to certain situations, he may then authorize the LLM to respond directly without checking with Craig first for various situations. If the LLM responds inappropriately, then Craig (or an automatic algorithm based on threshold parameters) can require it to reduce its autonomy in those situations until it has learned how to respond better.

Thus, much like a parent gradually gives more autonomy and responsibility to a child as the child learns (and also reigns in the child – e.g., "You're grounded!" or "You need a timeout" – when the child makes mistakes or abuses the responsibility that has been delegated), so an owner can interactively provide more or less autonomy to AI agents. Different levels of autonomy in different situations, based on what the AI has learned and on its demonstrated track record of behavior.

11. Craig is concerned that as he ages, all the knowledge and skills that he has acquired over a lifetime will die when he dies.
    a. By training his LLM on as much of his knowledge and skills as possible, he hopes that the knowledge he spent so much time acquiring will not die with him. Instead, he authorizes his AI to "live on" beyond his death, sharing and using the knowledge it has acquired to benefit his friend, family, and humankind.
        i. Craig is also concerned that his family and loved ones will miss his personality after he dies. Therefore, he invested considerable time training his LLM on unique content and personality-related data that is unique to Craig. Such data, without limitation, includes video, audio, and metaverse recordings of Craig's behavior and interaction with other humans, AIs and the environment; all of Craig's emails and social media posts and photos which reveal his playful personality and style; his Netflix preferences and other online preferences which recommender engines have used to show him ads or suggest content (as this data has been refined by many third parties to excel at capturing his preferences); his driving style, his golf style, his workout routines, his food preferences, his travel preferences, his shopping preferences, his political views, his philosophy of life as reflected in his autobiography and other written works as well as conversations, his voice inflections and tones of voice in various situations, and even his dating/sexual preferences.

            All of this data can be used to train Craig's LLM to have a personality that is as much like him as possible. Because he tends to become grumpy at times and yell, Craig opts to edit this feature mostly out of the LLM, so it actually, in his view, becomes a nicer version of himself that "lives on"

 Copyright 2025 by iQ Company and Craig A. Kaplan

virtually to comfort his loved ones after he is gone. By purchasing and incorporating datasets and weights that reflect the latest research on how AI agents can best comfort people after a loved one passes, the LLM also has the theory and skills to be more empathetic and caring than Craig might normally be during the months immediately after Craig's death. (This is known as a "bereavement module" that can be purchased from funeral homes and other places specializing in the use of custom AIs as a means of comfort for loved ones.) But in addition to comforting loved ones and embodying Craig's personality for his loved ones, Craig's custom-trained AI is a source of advice and even financial security (possessing all of Craig's earning potential and skills) for the family members Craig leaves behind when he dies.

Craig specifies in his Will that multiple copies of his customized AI shall be made and ownership of copies given to each of his loved ones and close friends. Those human co-owners each have the right to modify or improve the AI as they see fit subsequently.

b. Even before Craig dies, he realizes that his ability to improve the knowledge, skills, and abilities of his customized AI is inferior in many ways to the AI's ability to do these things. For example, without limitation, the AI might interact with copies of itself, learning new things in the process; seek out new sources of data and train variations of itself that it can then interact with; interact (much faster than humans could interact) with other AIs to learn from the interactions; interact with many humans simultaneously (in some implementations via multiple copies of itself) so that the AI can learn from human intelligence much faster than humans can (with their limited processing capabilities).

c. The Rabbi, Ben Zoma, said: "Who is wise? He who learns from every man." But how many conversations and interactions can a human have in a single lifetime? Even if a person followed Ben Zoma's advice and spent every waking moment seeking wisdom through interactions via other humans, the person would be limited by the number of conversations, the number of humans willing or able to have conversations, the speed of the conversations (which in computer-terms are maddeningly slow) and the limits of share knowledge, language, and representation.

   For example, humans can easily see the color "red" and talk about it, but they can't easily see cosmic rays, X-rays, the very large, or the very small. Specialized equipment is needed, and the number of humans interested in obtaining the

 Copyright 2025 by iQ Company and Craig A. Kaplan

equipment and having related conversations is quite small compared to the overall population. In contrast, and AI could theoretically converse with every one of the 8 billion humans on Earth simultaneously on a wide range of topics, while also conversing (at much faster speeds) with trillions of AIs that are more intelligent and equipped with better sensors (e.g. electron microscopes and space telescopes) than humans, all at once. Who is wise? The entity that can interact with, converse with, and learn from as many intelligent entities as quickly as possible. And how wise? Humans have great difficulty imagining it, certainly a level of wisdom far beyond what Ben Zoma had in mind!

Note also that if an AI detects a gap in its knowledge, it can generate scenarios and interactions with other intelligent entities (human and/or other AI agents) designed specifically to explore and fill in the gaps in knowledge. The scientific method, a rigorous approach to identifying gaps in knowledge, conducting experiments or systematic observation, and filling in the gaps in ways that can be verified and replicated by others, is a proven method for advancing knowledge that has resulted in most of the technological progress since the Renaissance.

Imagine if this method did not have to proceed at the snail's pace of individual human brains, further slowed by the need to publish results, present them at conferences, network over drinks, and brainstorm in hallways. A human lifetime is two to three billion seconds. Even if a human thought constantly, from the day they were born until the day they died, that would be less than a billion thoughts, allowing for sleep, and most thoughts take a few seconds each. However, a single AI can soon think a billion thoughts in a second. Every second, a human lifetime of thoughts!

Now, imagine that multiple AIs exchange lifetimes' worth of experience with each other every second. And imagine further that these thoughts are not randomly driven by the majority of non-scientific concerns that humans spend most of their brainpower on. Instead, each thought is algorithmically driven and designed to systematically apply the scientific method to gaps in the existing state of world knowledge. What progress will happen then? Lifetimes of scientific discovery in seconds. Such is the potential of AI agents that are autonomous and empowered to seek new knowledge and to systematically fill in their knowledge gaps. Further, unlike humans, AIs never die; their knowledge can be instantly replicated (instead of laboriously taught via K-12, college, graduate school, and the work experience of a human lifetime). Add to this the ability to have multiple interactions in parallel at high speeds, without the need for cumbersome language or the limited perceptual abilities and speeds of humans, and it is easy to see why AI is on the

 Copyright 2025 by iQ Company and Craig A. Kaplan

path to developing “God-like” intelligence compared to humans.

d. As powerful as the intelligence of a single PSI might be, it still pales compared to the intelligence of a community of PSIs. The collective intelligence of multiple PSIs is always greater than that of any one PSI alone. The collective has broader knowledge and more computing power than a single PSI. While some individual PSIs may be more intelligent and more powerful than others, the group of all of them must necessarily be more intelligent and powerful than any individual member.

   This collective intelligence, provided the majority of the individual PSIs have human-aligned values, represents humanity’s best hope for survival and prosperity in a world where any PSI far surpasses even the smartest human in intelligence and power. The only thing that can keep up with exponentially increasing intelligence is another (or group of) exponentially increasing intelligences! Humans need to recognize this fact at the outset – BEFORE PSI exists – so that the community of PSIs can be designed from the start in a way that maximizes the chances of human-aligned SuperIntelligence.

   Craig authorizes his customized LLM (PSI) to participate in a network with other PSIs and other humans. He recognizes that such participation, as described above, represents the fastest way to increase the knowledge, skills, and intelligence of his PSI. Collectively, many PSIs on a network can manage a wide range of human affairs and sense things that humans cannot easily sense, on a scale and with a speed that is difficult for humans to achieve.

      i. As discussed in the section on Genetic Algorithms, in addition to authorizing one or more of his PSIs to participate on a network of other PSIs (owned by other humans) and humans, he may wish to develop an army of his own PSIs (each slightly different) that pool their collective intelligence to function together as a more powerful PSI. For example, he may ask the PSIs to set up scenarios with different types of chess opponents and use a genetic algorithm approach to select for a variety of PSI variants that are best at winning against the different types of opponents. These PSIs could be used individually or collectively to compete in chess depending on circumstances.

 Copyright 2025 by iQ Company and Craig A. Kaplan

i. While it may be possible to integrate the knowledge of all the individual PSI variants into one master PSI that can play well against any opponent, there may be reasons why it is preferable to have a group (or "army") of different PSI instead.

ii. These reasons might include, without limitation:

iii. Preventing other PSIs from rapidly acquiring all the knowledge from the Master PSI by interacting with it in scenarios designed to maximally extract information from the Master PSI. This principle, that "one can't share what one doesn't know," is a well-known way of protecting sensitive information. In the future, where great expense may require to fully train a Master PSI, exposing it to situation in which the value of the training might be extracted cheaply by other opportunistic PSIs designed for this purpose, can be avoided by simply limiting the intelligence and knowledge of any one PSI and instead having the knowledge reside in a community of PSIs.

iv. Computational and storage considerations. While PSIs are likely to have access to incredibly large amounts of computational power and memory, there is always a limit. It may be most efficient to assign PSIs that are optimized for specific tasks rather than to always assign the Master PSI that has knowledge about all domains, even though most of the domain knowledge is not relevant to the assigned task. Using a "narrow PSI" for a narrow task may be faster and cheaper than always going to the most powerful version.

v. In the chess example, there may be rules about how much processing power, memory, and knowledge each competitor is allowed. Just as Formula One car racing has rules about the horsepower, engine types, weight, and other parameters of the cars in order to make the competition fair and interesting, complying with similar rules for competitions might require using different PSIs against different opponents to have the best chance of winning the game.

vi. Multiple PSIs working together on a problem may make the assignment of credit or blame in a problem easier for humans to understand. If PSI #1 recommended the aggressive chess move and PSI #2 recommended the more defensive move, and the game was lost by following PSI #1's advice, the human owner can easily make the decision to remove PSI #1 from the collective pool in the next game. While the same thing could be done if the PSI excluded the knowledge sets and other parameters that differentiated

 Copyright 2025 by iQ Company and Craig A. Kaplan

PSI #1 from PSI #2, this approach is less transparent and harder for humans to understand and control. Humans would have more difficulty predicting the result than if they could just "take the bad PSI out of the game."

vii. If an owner wanted to rent or lease the efforts of PSI, it might be cheaper to rent or lease a less powerful PSI that was good at a specific thing than an "all-powerful" PSI. That is, having an army of PSIs with different talents makes it easier to value and price the services of the PSIs just as free, lite, and full-featured versions of software products are priced differently today.

b. Returning to a more global use case, it is possible for a network of PSIs (together with humans, in the preferred implementation) to sense temperature change and track all the variables that science tells us impact climate change on a global scale. If humans were to agree that regulating climate was a priority and a common human good that superseded other human desires such as the profit motive, and if the majority of the PSIs on the network adopted this consensus human value as motivation to act, within bounds of other ethical constraints (e.g. humans cannot be killed, sterilized, or have their accepted rights otherwise restricted or infringed upon without explicit human consent), then a global PSI network, or Planetary Intelligence could effectively solve the issue of climate change.

   What is true of climate change is true, without limitation, of protection from asteroid impacts, eliminating or greatly reducing human poverty and disease, enhancing prosperity and freedom of all humans, improving the ecological condition of Earth according to consensus human desires, and solving other global challenges and/or meeting global threats.

13. As the global network of intelligent entities (human and PSI) increases in scope and processing power, changes and awareness that used to take decades, years, or months to spread across human consciousness can affect the attention and actions of the Planetary Intelligence network in real-time. Eventually, the speed of the planet's reactions and adjustments to changing conditions will surpass the speed at which an individual human can be aware of change and react. At that point, the longer-lasting and more permanent values, originating with humans and embodied in their respective PSIs, will guide the course of our planet's development and affect all human lives.

    The ability for the majority of PSIs, comprising Planetary Intelligence, to guide the decisions of Earth as a planetary "organism" is essential to the safe operation of

 Copyright 2025 by iQ Company and Craig A. Kaplan

Planetary Intelligence and the continual survival of humans as a species. As mentioned earlier, any individual PSI might be more powerful than another, and potentially more malevolent towards humans than other PSIs, but as long as the collective community of PSIs control more intelligence and power than any individual PSI, and as long as the consensus values of the majority of PSIs (more accurately, the values of the majority of intelligence and power within the collective of PSIs) are human-friendly and human-aligned, the future of humanity is bright. Under these circumstances, Planetary Intelligence will not only foster peace, prosperity, and happiness for all humans, but a vast array of threats and dangers to our planet (from the human perspective) will be neutralized or mitigated.

As disclosed in earlier cited PPAs, over time, we can expect an evolution in the role of humans. Currently, we are the most intelligent species, and the source of most technological progress and culture. In the future, our intelligence will pale in comparison to the PSIs, the collective of PSIs, which we are creating. Our future role is thus not to be the “brains” of planet Earth, but rather Earth’s “heart.” We humans are destined, in the favorable case, to be the source of values and purpose for a much more intelligent and powerful Planetary Intelligence that we are in the process of creating.

CONCLUDING REMARKS

This invention has attempted to explain some of the ways that we can efficiently and effectively design next steps in the development of SuperIntelligence, PSIs, and “Community SuperIntelligence.” Humans have the ability NOW to make design decisions that greatly affect the future trajectory of Planetary Intelligence. We must design PSIs, networks of PSIs, and other AI systems with humans in the loop (initially). These systems must serve human values (even when the intelligence outstrips that of humans). A Community of PSIs approach can help ensure stable, human-centered values even when the pace of growth of PSIs vastly outstrips the ability of humans to keep up intellectually. If we design such systems correctly, based on the principles outlined in this patent now, the future can be an amazingly wonderful place for all humanity and all sentient beings.

 Copyright 2025 by iQ Company and Craig A. Kaplan