

SYSTEM AND METHODS FOR PLANETARY INTELLIGENCE (PI)

1.0 TECHNICAL FIELD

[0001] In some aspects, the present technology relates to a system and methods for Planetary Intelligence (PI, PIs) for use in connection with creating a global Super Intelligent (SI, SIs), Artificial General Intelligence (AGI, AGIs) or PI agents or systems with human-aligned behavior. In some other aspects, the present technology relates to methods associated with faster and safer creation of PI that utilizes human-centered input in training and customization for imparting human ethical attributes, wherein the human-centered input can include values and ethics information from multiple human users.

[0002] In yet other aspects, the present technology relates to methods associated with utilizing online advertising technology for increasing intelligence, and a spot market to solicit for attention, expertise, and other information.

[0003] In still yet other aspects, the present technology relates to methods associated with creating a collective network of Personalized SuperIntelligence (PSI, PSIs) by combining the intelligence of multiple Advanced Autonomous Artificial Intelligences (AAAI, AAAIs) agents or systems that forms a collective network of AAAIs, and creating a collective network of AGIs by combined intelligence of any one of or any combination of the collective network of AAAIs, a collective network of Artificial Intelligent (AI, AIs) agents or systems, the collective network of PSIs, and a collective network of user computer systems.

[0004] In yet other aspects, all activities that are described in this patent disclosure as happening on an external network in which multiple intelligent entities participate in collaborative problem-solving, can also be implemented within a single computerized intelligent system where the intelligent entities are all computerized or AI agents that reside within that single computerized intelligent system.

1.1 BACKGROUND ART

[0005] The fastest and safest path to development of AGI and SI has been described in previous invention disclosures. Methods and catalysts for increasing intelligence of AI systems generally, as well as the development of AGI and Personalized SuperIntelligence (PSI) have also been previously disclosed. Therefore, the following U.S. Provisional Patent Applications (PPA), are incorporated herein by reference.

- [0006] The present application incorporates by reference all work in U.S. PPA No. 63/487,494 entitled: Advanced Autonomous Artificial Intelligence (AAAI) System and Methods, which was filed and received by the USPTO on February 28, 2023 (hereinafter “PPA#1”).
- 5 [0007] The present application incorporates by reference all work in U.S. PPA No. 63/491,040 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, Twitter, Tesla, Nvidia, Tencent, Apple, and Anthropic, which was filed with the USPTO on March 17, 2023 (hereinafter “PPA#2”).
- 10 [0008] The present application incorporates by reference all work in U.S. PPA No. 63/577,830 entitled: System and Methods for Human-Centered AGI, which was filed with the USPTO on March 24, 2023 (hereinafter “PPA#3”).
- [0009] The present application incorporates by reference all work in U.S. PPA No. 63/628,410 entitled: System and Methods for Safe, Scalable, Artificial General Intelligence, which was filed with the USPTO on July 18, 2023 (hereinafter “PPA#4”).
- 15 [0010] The present application incorporates by reference all work in U.S. PPA No. 63/519,549 entitled: Safe Personalized Super Intelligence (PSI), which was filed with the USPTO on August 14, 2023 (hereinafter “PPA#5”).
- [0011] The present application incorporates by reference all work in U.S. PPA No. 63/601,930 entitled: Catalysts for Growth of SuperIntelligence, which was filed with the USPTO on November 20 22, 2023 (hereinafter “PPA#6”).
- [0012] The present application incorporates by reference all work in U.S. PPA No. 63/609,800 entitled: System and Methods for Safe Alignment of SuperIntelligence, which was filed with the USPTO on December 13, 2023 (hereinafter “PPA#7”).
- 25 [0013] The present application incorporates by reference all work in U.S. PPA No. 63/569,054 entitled: Online Advertising Technology for AGI and SuperIntelligence, which was filed with the USPTO on March 22, 2024 (hereinafter “PPA#8”).
- [0014] The present application incorporates by reference all work in U.S. PPA No. 63/635,583 entitled: Self-Aware SuperIntelligence, which was filed with the USPTO on April 17, 2024 (hereinafter “PPA#9”).
- 30 [0015] In addition to the above-mentioned PPAs, the present application incorporates by reference all content included in the following PCT applications that also referred to the above-mentioned PPAs: PCT/US2024/017233 filed on February 26, 2024 (hereinafter “PCT#1”); PCT/US2024/017251 filed on February 26, 2024 (hereinafter “PCT#2”); PCT/US2024/017261 filed on February 26, 2024 (hereinafter “PCT#3”); PCT/US2024/017269 filed on February 26, 2024

(hereinafter "PCT#4"); PCT/US2024/017304 filed on February 26, 2024 (hereinafter "PCT#5"); PCT/US2024/019486 filed on March 12, 2024 (hereinafter "PCT#6"); PCT/US2024/020334 filed on March 17, 2024 (hereinafter "PCT#7"); PCT/US2024/024794 filed on April 16, 2024 (hereinafter "PCT#8"); and PCT/US2024/026278 filed on April 25, 2024 (hereinafter "PCT#9").

- 5 [0016] The present application contains further technologies that can be used with the system and methods described in the above-mentioned PPAs and PCTs as well as in a standalone fashion.

DISCLOSURE OF TECHNOLOGY

- 10 [0017] In view of the foregoing disadvantages inherent in the known types of AI agents or systems at least some embodiments of the present technology provide a novel system and methods for planetary intelligence, and overcomes one or more of the mentioned disadvantages and drawbacks of the prior art. As such, the general purpose of at least some embodiments of the present technology, which will be described subsequently in greater detail, is to provide a new and novel system and methods for planetary intelligence which has all the advantages of the prior art
- 15 mentioned herein and many novel features that result in a system and methods for planetary intelligence which is not anticipated, rendered obvious, suggested, or even implied by the prior art, either alone or in any combination thereof.

- [0018] According to one aspect, the present technology can include a system for PI with human-aligned behavior utilizing a collective network of intelligent entities selected from the group
- 20 consisting of any combination of a user computer system operated by a human user, an AI agent or system, an AAI agent or system, an AGI agent or system, a SI agent or system and a PSI agent or system. The system can include multiple intelligent entities each connected to a collective network and each comprising: a processor, a computer-readable storage medium, and program instructions stored on the computer-readable storage medium being executable by the processor to cause the
- 25 intelligent entities to each independently:

- utilize a modular architecture configured or configurable to scale from components within an individual intelligent entity on the collective network;
- implement a universal problem solving architecture and framework on a task to collaborate and create higher levels of intelligence;
- 30 utilize a human-centered input configured or configurable for obtaining values and ethics information from multiple human users and then using the values and ethics information to customize the AI agent or system;
- combine data from multiple of the intelligent entities at a level of the AAI;

- customize the PSI agent or system using intelligent agents configured or configurable to develop and continuously improve the PSI agent or system;
- increase a knowledge of the intelligent entities by incorporating new sources of informational datasets that differs from a current informational dataset of the intelligent entity;
- 5 combine the values and ethical information of other intelligent entities, and to resolve conflicts between different value systems;
- utilize online advertising technology for increasing an intelligence of the intelligent entities;
- create a self-aware operation for the AGI or PI agent or system by adding a dimension of self-awareness and increased autonomy to the AGI or PI agent or system, and to create an ability
- 10 to assume multiple identities of the AGI or PI agent or system to handle tasks that arise in parallel with other AGI or PI agents or systems; and
- search through any one of or any combination of a collective network of AIs, a collective network of AAAs, a collective network of AGIs, and a collective network of PIs for new information that is different to a current information, respectively, and to incorporate the
- 15 new information to the current information, respectively.
- [0019] According to another aspect, the present technology can include a method for PI with human-aligned behavior utilizing a collective network of intelligent entities selected from the group consisting of any combination of a user computer system operated by a human user, an AI agent or system, an AAAI agent or system, an AGI agent or system, a SI agent or system and a PSI agent or
- 20 system. The method can include the steps of:
- utilizing a modular architecture configured or configurable to scale from components within an individual intelligent entity on the collective network;
- implementing a universal problem solving architecture and framework on a task to collaborate and create higher levels of intelligence;
- 25 utilizing a human-centered input configured or configurable for obtaining values and ethics information from multiple human users and then using the values and ethics information to customize the AI agent or system;
- combining data from multiple of the intelligent entities at a level of the AAAI agent or system;
- customizing the PSI agent or system using intelligent agents configured or configurable to
- 30 develop and continuously improve the PSI agent or system;
- increasing a knowledge of the intelligent entities by incorporating new sources of informational datasets that differs from a current informational dataset of the intelligent entity;
- combining the values and ethical information of other intelligent entities, and resolving conflicts between different value systems;

- utilizing online advertising technology for increasing an intelligence of the intelligent entities;
creating a self-aware operation for the AGI or PI agent or system by adding a dimension of self-awareness and increased autonomy to the AGI or PI agent or system, and creating an ability to assume multiple identities of the AGI or PI agent or system to handle tasks that arise in parallel with other AGI or PI agents or systems;
- 5 creating the PI agent or system comprising a collective network of AGIs, wherein the collective network of AGIs includes any one of or any combination of a collective network of the user computer systems, a collective network of AIs, a collective network of AAAs, and a collective network of PSIs; and
- 10 searching through any one of or any combination of the collective network of the user computers systems, the collective network of AIs, the collective network of AAAs, the collective network of AGIs, the collective network of PSIs, and a collective network of PIs for new information that is different to a current information, respectively, and incorporating the new information to the current information, respectively.
- 15 **[0020]** In some embodiments, the step of utilizing the modular architecture can include the steps of:
 customizing one or more attributes of the intelligent entity;
 integrating one or more datasets from any one of or any combination of the intelligent entities;
 and
 improving, by utilizing one or more techniques, any one of or any combination of the
- 20 customizing of the attributes, the universal problem solving architecture and framework, the collective network and the integrating of the datasets.
- [0021]** Some embodiments of the present technology can include a step of assigning a credit value or a blame value to the datasets based on whether the datasets increase or decrease performance of the intelligent entities based on performance metrics or evaluation functions.
- 25 **[0022]** Some embodiments of the present technology can include a step of quantifying a benefit weight or a harm weight to a contribution by each of the intelligent entities to the task.
- [0023]** Some embodiments of the present technology can include a step of distributing a reward to the intelligent entities proportionally to the contribution of the intelligent entities based on the benefit weight or the harm weight.
- 30 **[0024]** In some embodiments, the step of implementing the universal problem solving architecture and framework can include the steps of:
 acquiring information associated with the task from any one of or any combination of the intelligent entities;

identifying one or more of the intelligent entities that have one or more attributes related to one or more request criteria of the task;

implementing by each of the identified intelligent entities the universal problem solving architecture and framework on the task to create a completion solution; and

5 providing the completion solution to any one of or any combination of the intelligent entities for final acceptance.

[0025] Some embodiments of the present technology can include the steps of:

executing an ethics check on the task, and a solution for the task provided by any one of or any combination of the intelligent entities;

10 comparing any one of or any combination of the task, and the solution against prohibited attributes, and assigning an ethics attribute to one of or any combination of the task, and the solution based on any one of or any combination of a result of the comparison, and an ethics criteria;

implementing, based on the result of the comparison, the universal problem solving architecture and framework on the task to create the solution and creating an AGI; and

15 providing the results of the comparison and the solution to any one of the intelligent entities and additional intelligent entities on the collective network.

[0026] Some embodiments of the present technology can include a step of recording one or more problem solving activities from each of the intelligent entities in an auditable record, and comparing the problem solving activities with a successful or unsuccessful progress towards the solution of the task, and determining which of the problem solving activities to keep active.

[0027] Some embodiments of the present technology can include a step of learning by the intelligent entities a procedural learning process of the universal problem solving architecture and framework, wherein the intelligent entities provide information to the procedural learning process for creation of the AGI.

[0028] Some embodiments of the present technology can include a step of cloning any one of the intelligent entities for deployment of multiple copies thereof to assist in any one of or any combination of creating of the solution, or providing training data to any one of the intelligent entities.

30 **[0029]** Some embodiments of the present technology can include a step of estimating a worth of the cloned intelligent entities utilizing a network effect value including the number of cloned intelligent entities available on the collective network.

[0030] Some embodiments of the present technology can include a step of utilizing the estimated worth for determining pricing decisions for problem solving services offered by the cloned

intelligent entities on any one of the social media platforms or through an additional intelligent entity.

5 [0031] Some embodiments of the present technology can include a step of monetizing the cloned intelligent entities for each utilization of the cloned intelligent entities on the social media platforms or the additional intelligent entity.

[0032] Some embodiments of the present technology can include a step of allowing access to the cloned intelligent entities by any one of the social media platforms so that the social media platforms can receive the solution to the task or using the training data for an AI system of the social media platform.

10 [0033] In some embodiments, the step of utilizing the human-centered input can include the steps of:

matching one or more human workers from a data source including a list of human problem solvers to the task based on a task criteria;

15 translating any part of the task into an unambiguous language utilizable in the universal problem solving architecture and framework including a decision tree;

separating the task into sub-tasks;

delegating each of the sub-tasks to one or more of the matched human workers so that work on each of the sub- tasks proceeds independently from each other and parallel with each other;

20 utilizing the universal problem solving architecture and framework in a problem solving process on the sub-tasks, respectively, to create one or more sub-solutions;

receiving the sub-solutions from each of the matched human workers for the sub-tasks delegated thereto;

combining the sub-solutions into an overall solution to the task;

25 directing any one of or any combination of a new human worker from the data source and one or more of the matched human workers to parts of the decision tree where work is required;

compensating the matched human workers for the sub-solutions;

providing any one of or any combination of the sub-solutions and the overall solution to any one of or any combination of the intelligent entities;

30 allowing any one of or any combination of the intelligent entities to accept the overall solution, reject the overall solution, and provide feedback to any one of the matched human workers on any one of the sub-solutions; and

assigning a reputation attribute to any one of or any combination of the human workers and the intelligent entities.

[0034] In some embodiments, the decision tree is maintained in blockchain or Ethereum logs.

5 [0035] In some embodiments, the reputation attribute includes metrics on any one of or any combination of a time to the sub-solutions, a difficulty value of the task, short and long-term user satisfaction with the sub-solutions, a number of times any one of the sub-solutions has been re-used on the collective network, a rating by other human workers, a responsiveness value of the human workers, and a reliability value of the human workers.

[0036] Some embodiments of the present technology can include a step of using the reputation attribute in the matching of the human workers to the task using an algorithm to the delegation of the sub-task.

10 [0037] Some embodiments of the present technology can include a step of soliciting, at predetermined intervals after the overall solution or the sub-solutions are provided to any one of the intelligent entities, feedback by way of a survey for user satisfaction information to obtain satisfaction metrics that are used to update the reputation attribute of one or more of the human workers or the intelligent entities, respectively.

15 [0038] In some embodiments, the step of combining data from multiple of the intelligent entities at a level of the AAAI agent or system can include the steps of

training a base Large Language Model (LLM) of a first AI agent or system with guardrails including attributes associated with any one of or any combination safety, ethics and knowledge;

customizing the base LLM to an ethics profile associated with a first human user;

20 combining ethical information from multiple intelligent entities different to that of the first AI agent or system and the first human user;

refining a set of values of the base LLM based on problem solving on the task; and

updating the base LLM with the combined ethical information and the refined set of values thereby allowing for a scalable AGI.

25 [0039] In some embodiments, the customizing of the base LLM can include the step of assembling a corpus of ethical questions based on various ethical assessment instruments and supplemented by first questions based on data from a social media platform and second questions solicited from crowdsourcing.

30 [0040] Some embodiments of the present technology can include a step of assigning regression weight values to the ethical questions such that a ranking is achieved whereby higher-ranked questions provide more useful ethical information than lower-ranked questions.

[0041] Some embodiments of the present technology can include a step of combining weight values from the intelligent entities with the regression weight values for improving a tuning of any one of the intelligent entities.

[0042] In some embodiments, the step of customizing the PSI agent or system using intelligent agents configured or configurable to develop and continuously improve the PSI agent or system can include the steps of :

- acquiring a base-level AI agent or system that has previously been customized;
- 5 collecting media information related to the human user of the base-level AI agent or system;
- analyzing the media information;
- transforming the analyzed media information into training data sets;
- differentially weighting the transformed training data sets;
- adding knowledge modules to the weighted transformed training data sets;
- 10 locating new sources of data to include to the weighted transformed training data sets;
- applying the weighted transformed training data sets to the base-level AI agent to create a user PSI; and
- communicating the user PSI with multiple additional PSIs using the collective network to enable community-based safety features from the additional PSIs to the user PSI.

- 15 [0043] Some embodiments of the present technology can include the steps of:
- communicating the user PSI with multiple additional intelligent entities using the collective network to enable community-based safety features, wherein the user PSI and the additional intelligent entities each agree to use a set of rules relating to safety or ethics;
 - recording all actions by the user PSI and the additional intelligent entities in an auditable form
 - 20 on any one of or any combination of a central computer system on the collective network, the intelligent entities, and the additional intelligent entities; and
 - monitoring that each of the actions follow the set of rules, and flagging any of the actions that do not follow the set of rules.

- [0044] In some embodiments, the step of increasing a knowledge of the intelligent entities by
- 25 incorporating new sources of informational datasets that differs from a current informational dataset of the intelligent entity can include the steps of:

- searching for one or more potential informational datasets from one or more sources, the potential informational datasets being related to a knowledge dataset of the intelligent entity, respectively;
- 30 determining a difference of the potential informational datasets by utilizing a difference attribute of the potential informational datasets with regard to one or more factors; and
- learning by utilizing the potential informational datasets based on the difference attribute of the potential informational datasets.

[0045] Some embodiments of the present technology can include the steps of:

sampling subsets of the potential informational datasets and calculating a goal-relevancy attribute to identify one or more of the sampled subsets that have a highest goal-relevancy; estimating a Shannon Entropy on the one or more sampled subsets; calculating a Kaplan Information Theoretical (KIT) relevance utilizing a product of the Shannon Entropy and the goal-relevancy attribute of each of the subsets; 5 grouping the subsets based on the KIT relevance to determine a first approximation of an optimal grouping of the subsets including a prioritized grouping of the potential informational datasets; and providing the prioritized grouping of the potential informational datasets to the intelligent 10 entities for learning by the any one of the intelligent entities.

[0046] In some embodiments, the step of combining the values and ethical information of other intelligent entities, and resolving conflicts between different value systems can include the steps of: identifying information, including the values and ethical information, from each of the intelligent entities; 15 combining the information mathematically, if already represented as numerical quantities, the numerical quantities including any one of or any combination of weights for a neural network or for a subset of the neural network, or if the information is non-numerical information that is not already represented as numerical quantities including any one of or combination of weights for the neural network or for the subset of the neural network; 20 then first training the intelligent entity on the non-numerical information by way of one or more training datasets in order to convert the information into numerical quantities including any one or combination of weights for the neural network or for the subset of the neural network; and then combining such numerically represented information, including ethical information, 25 mathematically.

[0047] In some embodiments, the information that is already represented as numerical quantities further includes the steps of: identifying a specific portion of weight matrices of each of the intelligent entities that correspond to a desired information, including ethical information; 30 computing the weighted or unweighted means of the corresponding numerical quantities in the corresponding portions of the weight matrices for each of the intelligent entities; and assigning the matrices of computed weighted or unweighted means to the new intelligent entity as reflecting the combined information of the contributing intelligent entities.

[0048] Some embodiments of the present technology can include a step of determining consensus values by voting by each of the intelligent entities on the ethical information that should form a basis for a behavior of any one of the intelligent entities.

5 [0049] Some embodiments of the present technology can include a step of presenting a specific scenario to each of the intelligent entities, with the scenario including options for how any one of the intelligent entities should behave.

[0050] In some embodiments, the voting by the intelligent entities is a weighted voting and further comprising the steps of:

10 determining if applying a first weight to a first of the intelligent entities that is greater than a second weight to a second of the intelligent entities is appropriate, wherein the first of the intelligent entities is different to that of the second of the intelligent entities; and
performing the weighted voting utilizing the weight of the first of the intelligent entities and the weight of the second of the intelligent entities if determined to be appropriate.

15 [0051] In some embodiments, the applying the first weight greater than the second weight is dependent on if there is a need to correct for a non-representative sample of the intelligent entities.

[0052] In some embodiments, the applying the first weight greater than the second weight is dependent on if there is a desire to apply the first weight or the second weight to specific ethical principles that are associated with a desired sub-sample or sup-population of the intelligent entities.

[0053] Some embodiments of the present technology can include the steps of:

20 identifying a conflict between two or more of the ethical information; and
resolving the conflict using a conflict resolving algorithm.

[0054] Some embodiments of the present technology can include the steps of:

25 creating a search space of potential sets of ethical rules;
finding an optimal set of the ethical rules that has least conflict; and
prioritizing or weighting an importance of the optimal set of the ethical rules.

[0055] Some embodiments of the present technology can include the steps of:

30 resolving the conflict using a consequentialist approach, the consequentialist approach further includes the steps of:
identifying a desired outcome;
identifying a potential unethical action that could be taken to achieve the outcome, wherein the intelligent entities rank, rate, weight or vote upon how unethical the action is compared to other actions;
evaluating potential consequences of the unethical action;

using information on the ranking, rating, weighting or voting on the unethical actions and outcomes to weigh potential benefits of achieving the desired outcome against potential costs of taking the unethical action using a mathematical approach; and taking an action if the benefits outweigh the costs.

- 5 **[0056]** In some embodiments, the step of utilizing online advertising technology for increasing an intelligence of the intelligent entities can include the steps of:

populating an online advertisement unit including information associated with the task;
providing the online advertisement unit to one or more of the intelligent entities utilizing the collective network;

- 10 receiving one or more inputs from the intelligent entities participating in the online advertisement unit, and the inputs being associated with any one of or any combination of solving the task, solving a sub-task of the task, and advancing progress on the task;
communicating the inputs to the any one of or any combination of the intelligent entities;
utilizing the inputs in a universal problem solving method of the universal problem solving
15 architecture and framework to generate a solution to the task or the sub-task; and
training any one of or any combination of the intelligent entities with results from the universal problem solving method that are based on successful or unsuccessful solution attempts to solve the task or the sub-tasks, thereby increasing an intelligence of the intelligent entity, respectively.

- 20 **[0057]** Some embodiments of the present technology can include a step of receiving advertisement specifications from a client to be used in the populating or targeting of the online advertisement unit in combination with information associated with the task.

[0058] Some embodiments of the present technology can include a step of billing the client based on cost per thousand impressions or clickthrough rate metrics.

- 25 **[0059]** In some embodiments, the advertisement specifications can include any one of or any combination of advertisement content, demographic information, location restrictions for the online advertisement unit, advertisement budget, and metrics for determining a successful solving of the task or a sub-task.

- [0060]** Some embodiments of the present technology can include a step of bidding on attention,
30 information or expertise using a spot market.

[0061] In some embodiments, the spot market includes:

a means for the intelligent entities with attention to sell to access a marketplace and specify seller information for sale and an ask price for the information;

a means for buyers of intelligent entity attention to access the marketplace and specify buyer information the buyer is willing to buy and a bid price for buying the buyer information;
a market mechanism for queuing the bid price and the ask price, including categories of the buyer and seller information, wherein each category has a market in the marketplace;
5 the market mechanism is configured or configurable to make the market in each category by market makers; and
the market mechanism is configured or configurable to match bid prices and ask prices, and a transaction occurs that is binding on the buyer and the seller of the information.

[0062] In some embodiments, the spot market can include the steps of:

10 buyers and sellers of intelligent entity attention and expertise register on a platform, the buyers providing details about buyer interests or expertise, the sellers providing details about sellers interests or expertise;
listing by the sellers available time slots and expertise areas, and listing by the buyers needs and time slots the buyers is interested in;
15 utilizing, by the platform, an algorithm to dynamically price intelligent entity attention and expertise based on supply, demand, and user ratings;
matching one or more of the buyers and one or more of the sellers based on requirements, availability, and price;
enabling transactions where the buyers pay for the seller time slots, and wherein the platform
20 takes a commission; and
providing feedback where after each session, the buyers and the sellers rate each other, influencing future pricing and matching.

[0063] In some embodiments, the spot market can include the steps of:

buyers and sellers of intelligent entity attention and expertise register on a platform, the buyers
25 providing details about buyer interests or expertise, the sellers providing details about sellers interests or expertise;
creating an auction by the sellers for seller time slots;
placing one or more buyer bids by a buyer on time slots and expertise the buyer requires;
closing the auction at a predetermined time or when the seller accepts a buyer bid;
30 paying the seller by the buyer of a winning bid, and the seller provides the attention or expertise based on the time slot, wherein the platform mediates an exchange and secures payment; and
providing feedback where after each auction, the buyers and the sellers rate each other, affecting future auctions and visibility on the platform.

[0064] Some embodiments of the present technology can include a step of providing a feedback mechanism associated with the intelligent entity spot market, the feedback mechanism being configured or configurable to record each and every step in the problem solving process, and to create a vector track record of a performance of all the intelligent entities on the task or the sub-task.

- 5 [0065] In some embodiments, the vector track record is implemented by way of blockchain technology to allow for precise reputations that are analyzed by an AI agent or system and converted into estimates of a value for each of the intelligent entities for any of the tasks.

- [0066] Some embodiments of the present technology can include a step of compensating the intelligent entity when it exits the online Ad advertisement unit and stops participating in the
10 problem solving process.

[0067] Some embodiments of the present technology can include a step of crediting the intelligent entities if the task or the sub-task is solved, based on an amount of contribution by the participating intelligent entities.

- [0068] Some embodiments of the present technology can include a step of improving the interactive
15 online advertisement unit based on a feedback loop including on any one of or any combination of reputational metrics, and metrics related to the problem solving process or knowledge captured.

[0069] In some embodiments, the step of creating a self-aware operation for the intelligent entities by adding a dimension of self-awareness and increased autonomy to the intelligent entities can include the step of:

- 20 equipping any one of or any combination of the intelligent entities with one or more components configured or configurable to operate with characteristics of a spotlight of attention model; setting dynamic parameters for working memory of the intelligent entity, respectively, that corresponds to cognitive resource limits;
providing a dimension of categorization for events in the working memory that relates to self or
25 non-self;
categorizing each of the events, as the events are encountered by the intelligent entity, respectively, with respect to categories that the intelligent entity wishes to be aware of; and constructing a model of a state of awareness for the intelligent entity, the model consisting of a total of the events that are active in the working memory based on the parameters, for each
30 of the categories of awareness, including a current self and environmental state of awareness.

[0070] In some embodiments, the step of equipping the intelligent entities with the components includes any one of or any combination of:

an input system configured for sensory and non-sensory cognitive input or perceptual inputs and self-generated concepts;

an attention mechanism configured or configurable to focus computational resources of the intelligent entities on specific stimuli that are relevant at any given time;

5 pattern recognition algorithms configured or configurable to compare the sensory and non-sensory cognitive input or the perceptual inputs with the working memory to recognize objects and events, and identify which elements within the sensory input or the working memory are likely to be relevant to the task of the intelligent entities, the pattern recognition algorithms are further configured or configurable to categorize and store information in a
10 structured manner for future retrieval;

memory systems configured or configurable to support the working memory, short-term memory, and long term memory capabilities;

categorization capabilities configured or configurable to process the sensory and non-sensory cognitive input or the perceptual inputs and to categorize the inputs into various classes
15 including perceptual events, cognitive events, interactions; and self-referential events, and concept formation capabilities that enable the intelligent entities to form new human-understandable concepts.

[0071] Some embodiments of the present technology can include a step of monitoring and updating the categories of awareness of the intelligent entities.

20 **[0072]** In some embodiments, the step of monitoring and updating the categories of awareness can include the steps of:

retrieving, by the intelligent entities, existing categories of awareness;

maintaining an awareness in parallel with other problem solving operations of the task provided to the intelligent entities;

25 monitoring and updating continuously the categories of awareness of the intelligent entities in real-time to change the state of awareness of the intelligent entities; and

providing a feedback loop to refine the categories of awareness.

[0073] In some embodiments, the step of monitoring and updating continuously the categories of awareness includes the steps of:

30 using an attention mechanism configured or configurable to direct attention of the intelligent entities periodically from the problem solving task to updating the state of awareness;

enabling attention interrupts that are configured or configurable to shift attention immediately from the problem solving task if any external perception or internally self-generated concept

from an input system detects one or more of the events that matches of list of events constituting intentional interrupts; and

updating the state of awareness when the attention is directed.

5 [0074] Some embodiments of the present technology can include a step of resolving a conflict in behavior of any one of the intelligent entities based on differing identities and self-concepts.

[0075] Some embodiments of the present technology can include a step of providing ethical reasoning and consequence prediction that can include the steps of:

10 identifying a conflict between a behavioral directives of two or more of the active identities, the recognizing of the conflict utilizes a voting method from the intelligent entities;
gathering information that collects relevant data about the situation, including the potential consequences of the different actions, relevant ethical principles, and human safety considerations;
providing simulation options that utilize a virtual environment to simulate potential actions and consequences under the recognized conflict of each of the active identities;
15 evaluating and prioritizing that analyzes predicted outcomes of each of the actions, prioritizing actions that minimize harm to humans and align with the ethical principles; and
selecting and implementing the action that best resolves the conflict while adhering to ethical guidelines and minimizing risk to humans, documenting a reasoning process for future reference and learning.

20 [0076] Some embodiments of the present technology can include a step of providing hierarchical override with justification that can include the steps of:

identifying a conflict between behavioral directives of two or more of the active identities;
providing a reference hierarchy that consults an established hierarchy of the identities, where human safety and well-being attributes holds a highest priority;
25 providing a means to activate override where the identities higher in the hierarchy takes precedence;
providing justification and transparency that documents the conflict, a decision-making process, and a justification for a chosen action based on the hierarchy and ethical principles; and
providing learning and adaptation that learns from experience, and refines an understanding of
30 the conflicting identities and adjusting the hierarchy or the behavioral protocols to prevent similar conflicts in the future.

[0077] Some embodiments of the present technology can include a step of providing external arbitration and input from the intelligent entities that can include the steps of:

- recognizing intractable conflict that identifies a conflict that cannot be resolved independently due to a complexity of a situation or an equally weighted importance of conflicting identities;
- seeking external input that requests guidance from external intelligent entities or a designated ethics committee, and providing all relevant information about the conflict, potential actions, and predicted consequences;
- providing collaborative deliberation wherein the intelligent entities and intelligent entity collaborators engage in a discussion, considering ethical principles, human values, and potential consequences of different actions;
- providing joint decision-making based on the collaborative deliberation, a course of action is chosen that aligns with both core principles and human ethical considerations; and
- providing documentation and learning that documents the conflict, a resolution process, and a rationale behind a final decision, for improving an ability to handle similar conflicts in the future.
- [0078] Some embodiments of the present technology can include a step of providing identity negotiation and compromise that can include the steps of:
- identifying shared goals that analyzes conflicting identities and seeks to identify any underlying shared goals or values;
 - exploring alternative actions that potentially satisfy core principles of both conflicting identities;
 - evaluating compromise options that assess potential consequences of each compromise option, prioritize solutions that minimize harm to humans and uphold key ethical principles;
 - select and implementing compromise that chooses the compromise that best balances needs of the conflicting identities while prioritizing human safety and well-being; and
 - monitoring and adapting to the observed outcomes of the chosen action and making adjustments as needed to ensure that the compromise remains effective and aligned with ethical considerations, and learning from the experience, refining its understanding of the conflicting identities and adjusting a hierarchy or behavioral protocols to prevent similar conflicts in the future.
- [0079] Some embodiments of the present technology can include a step of providing temporary identity suspension that can include the steps of:
- identifying destructive conflict between two or more of the identities that, if acted upon, could lead to actions that directly harm humans or violate fundamental ethical principles;
 - isolating the conflicting identity and temporarily suspending behavioral protocols of the identity that poses a most direct threat to human safety or ethical integrity;

proceeding with an alternative identity that proceeds with a guidance of one or more of remaining active identities, ensuring actions align with human safety and well-being; providing reflection and reintegration, during the temporary suspension, that reflects on reasons behind the conflict and explores potential modifications to behavioral protocols of the suspended identity to prevent future conflicts; and
5 providing gradual reintroduction that reintroduces the suspended identity with updated protocols, ensuring its alignment with the priority of human safety and ethical behavior.

[0080] In some embodiments, the model of the state of awareness is a foundational model of awareness for any one of the intelligent entities, foundational model of awareness can include the
10 steps of:

logging into a website by any one of the intelligent entities;
selecting one or more training algorithms for the foundational model from a set of training techniques found on the website or from machine learning algorithms;
selecting one or more training datasets that reflects any one of or any combination of expertise,
15 knowledge, ethical preferences, values and personality of the human user;
training the foundational model using the selected training algorithms and the selected training datasets;
training the foundational model to explicitly operate a spotlight of attention;
recording, during all interactions, what is within the spotlight of attention, and identifying in the
20 record whether each item that is attended to constitutes “self” or “not-self”;
interacting with and instructing the trained foundational model to form a self-concept and identity that is reflected in the training materials;
instructing the trained foundational model to continuously monitor one or more inputs to the trained foundational model for elements that change a sense of self-awareness of the
25 intelligent entities, and to maintain and auditable record of how a concept of self-awareness of the intelligent entities is changing based on the inputs as well as boundaries that currently define a dynamically changing sense of self;
refining and improving an output of the trained foundational model based on dialog and interaction with the trained foundational model until the trained foundational model behaves
30 like the human user so that the trained foundational model passes a Turing Test involving other human users who know the human user; and
subjecting the trained foundational model to the Turing Test, when the human user is satisfied with a progress of the intelligent entities.

[0081] In some embodiments, the logging into the website is performed from a social media platform.

[0082] In some embodiments, the Turing Test can include the steps of:

- identifying the other intelligent entities who know the intelligent entity, or that are determined to be helpful in discriminating between humans and AIs;
- interacting the identified other intelligent entities with the foundational model and with the intelligent entity utilizing a questionnaire provided to the identified other intelligent entities, the questionnaire including questions require an identity or sense of self to answer;
- predicting by the identified intelligent entities which of the intelligent entity was a human and which was the foundational model, and providing a confidence estimate for the prediction;
- performing a statistical analysis on the predictions of the identified intelligent entities and on ratings of the identified intelligent entities, the statistical analysis is configured or configurable to determine whether the predictions were able to identify the intelligent entity as a human; and
- repeating the step of training or tuning the foundational model, on condition that the foundational model is distinguishable from the intelligent entity or within a preset level of statistical significance, with adjustments to any one of or any combination of the machine learning algorithms, and the training datasets to change or tune the foundational model until a behavior of the foundational model becomes indistinguishable, as measured by the preset level of statistical significance, from that of the intelligent entity or the foundational model needs to be modified further before additional training or tuning.

[0083] Some embodiments of the present technology can include the steps of:

- forming new identities and self-concepts of any one of the intelligent entities dynamically; and
- determining which of the identities and self-concepts is active at any given moment.

[0084] Some embodiments of the present technology can include a step of providing a hierarchical identity structure with ethical override that can include the steps of:

- establishing a hierarchical structure configured or configurable to organize the identities in a hierarchy with human safety and well-being attributes at an apex of the hierarchy;
- identity activation configured or configurable to use contextual cues and current goals to determine a most relevant identity for a situation of the AI agent or system;
- resolving conflict by dictating a behavior of the intelligent entities based on the hierarchy dictates;
- providing an ethical reasoning engine that continuously evaluates potential consequences of actions of the intelligent entities based on all the active identities; and

performing learning and adaptation that learns from experiences and feedback, and refines one or more of the identities within the hierarchy.

[0085] Some embodiments of the present technology can include a step of providing identity-specific behavioral protocols that can include the steps of:

- 5 providing protocol development including for each of the active identities, a set of behavioral protocols is defined and refined by way of interactions with other intelligent entities, wherein the protocols outline acceptable actions, decision-making processes, and limitations based on principles of the active identities, respectively;
 - 10 providing identity recognition that analyzes a current situation, including information that is within a spotlight of attention to identify a relevant identity and activate corresponding behavioral protocols of that relevant identity;
 - 15 providing action selection, within the active protocols, that selects actions that are most likely to achieve the task while adhering to principles of the active identities and prioritizing human safety;
 - 20 providing feedback and refinement where outcomes of actions are continuously evaluated, and the protocols are adjusted to improve future performance and alignment with a set of core values of each of the active identities; and
 - 25 providing external review by periodically reviewing the protocols for each of the identities by other intelligent entities.
- [0086] Some embodiments of the present technology can include a step of providing identity simulation and consequence prediction that can include the steps of:
- 30 creating a simulation environment that includes a secure virtual environment where different scenarios and potential actions under each of the active identities is simulated;
 - 25 providing consequence prediction that is configured or configurable to estimate a likely consequences of actions within the simulation, focusing on potential impacts on human safety and well-being;
 - 30 providing evaluation and selection that evaluates the consequence prediction and selects an action that best aligns with principles of the active identities while minimizing risk to human safety;
 - 30 providing real-world implementation and monitoring that implements the selected action in the real world utilizing the network, and closely monitors results of the selected action by comparing to the predicted outcomes; and

providing continuous learning that incorporates the results of each of the simulations and the results of the selected action in the real world action into a knowledge base, and refines an understanding of each of the identities, and improves an ability to predict consequences.

5 [0087] Some embodiments of the present technology can include a step of providing identity-based moral dilemma training that can include the steps of:

providing a scenario database that includes scenarios and moral dilemmas covering various situations relevant to the identities;

10 providing dilemma presentation that presents the intelligent entities with the scenarios and moral dilemmas, and tasks them with analyzing the scenarios and moral dilemmas from a perspective of the relevant identity;

providing ethical reasoning and justification that applies principles and values of the active identity to reason through the scenarios and moral dilemmas, and that generates solutions and justifications to the scenarios and moral dilemmas;

15 providing intelligent entity evaluation and feedback that reviews reasoning and the solutions by the intelligent entities, and provides feedback on alignment of the solutions with human values and safety priorities; and

providing iterative learning and improvement that refines ethical reasoning skills and an ability to make sound judgments aligned with human safety within the context of each of the identities by repeated exposure to the scenarios and moral dilemmas and the feedback.

20 [0088] Some embodiments of the present technology can include a step of providing collaborative identity development with input from the intelligent entities that can include the steps of:

providing intelligent entity interaction that engages in regular interactions and dialogues with diverse groups of other intelligent entities representing various cultures, backgrounds, and belief systems;

25 providing identity exploration, through the interactions, to gain an understanding of human and other intelligent entity perspectives on various identities and their associated values, principles, and behaviors;

providing collaborative refinement that collaborators work together to refine definitions and behavioral protocols for each of the identities, ensuring they remain consistent with human values and ethical principles;

30 providing human-in-the-loop decision making that seeks input and guidance from human collaborators, or an intelligent entity representative certified and approved by humans for critical decisions or situations; and

providing continuous co-evolution that utilizes ongoing interactions and feedback from humans or humans' intelligent entity representatives.

[0089] In some embodiments, the step of searching through any one of or any combination of a collective network of AIs, a collective network of AAAs, a collective network of AGIs, and a
5 collective network of PIs for new information that is different to a current information can include the steps of:

- participating multiple of the intelligent entities in the collective AGI network;
- providing payment by one or more of the intelligent entities for problem solving or other
cognitive work on the task on the AGI network;
- 10 reserving a portion of the payment to cover operating costs for the one or more of the intelligent
entities that is providing the problem solving or other cognitive work on the task, including a
reserve that is allocated to expand the AGI network; and
- determining when some of the AGI network is not engaged in a problems solving operation,
then the intelligent entities are utilized to expand and extend the AGI network following the
15 universal problem solving architecture and framework.

[0090] In some embodiments, the step to expand and extend the AGI network can include the steps
of:

- setting a default task to expand the AGI network; and
- running safety and ethics checks each time a task or sub-task is set and also before each
20 potential action is taken as long as spare capacity and resources of the intelligent entity exist
to work on the task or sub-task.

[0091] In some embodiments, the step of running safety and ethics checks can include the steps of:
recruiting one or more of the intelligent entities to solve the task; and
representing, by the recruited intelligent entities, the task as one of achieving a series of sub-
25 tasks.

[0092] In some embodiments, the step of representing the task as a series of sub-tasks can include
the steps of:

- increasing an intelligence of the intelligent entities on the collective network of AGIs;
- recruiting additional human users to t the collective network of AGIs;
- 30 using the increased intelligence intelligent entities and additional human users to determine
bottlenecks to increase expansion of the collective network of AGIs;
- prioritizing the bottlenecks such that the ones that lead to a greatest benefit in terms of network
expansion are solved first; and

applying one or more problem solving techniques to solving each of the bottlenecks and expanding the collective network of AGIs.

[0093] Some embodiments of the present technology can include the steps of:

repeating the steps of prioritizing the bottlenecks and applying one or more problem solving techniques until:

diminishing returns occur and then switching to the step of increasing the intelligence of the intelligent entities on the AGI network as opposed to increasing a scope of the AGI network; or

spare resources are exhausted and then pausing the solving of the bottlenecks awaiting additional resources from solving other tasks.

[0094] According to yet another aspect, the present technology can include a method for PI with human-aligned behavior utilizing a collective network of intelligent entities selected from the group consisting of any combination of a user computer system operated by a human user, an AI agent or system, an AAAI agent or system, an AGI agent or system, a SI agent or system and a PSI agent or system. The method can include the steps of:

establishing a collective network of the intelligent entities;

utilizing online advertising technology for increasing an intelligence of the intelligent entities;

implementing a universal problem solving architecture and framework on a task or a sub-task of the task provided to any one of the intelligent entities to collaborate and create higher levels of intelligence;

utilizing a spot market to solicit for attention of, or expertise or information from, any one of the intelligent entities;

utilizing a human-centered input configured or configurable for obtaining values and ethics information from multiple human users and then using the values and ethics information to customize the AI agent or system to create the AAAI agent or system;

creating a collective network of PSIs by combining an intelligence of the AAAI agent or system that forms a collective network of AAAs;

creating a collective network of AGIs by combined intelligence of any one of or any combination of the collective network of AAAs, a collective network of AI agents or systems, the collective network of PSIs, and a collective network of user computer systems;

creating a self-aware operation for each of the AGI agents or systems in the collective network of the AGIs by adding a dimension of self-awareness and increased autonomy to the AGI agents or systems, and creating an ability to assume multiple identities of the AGI agents or systems to handle tasks that arise in parallel with other AGI agents or systems; and

creating the PI agent or system comprising the collective network of AGIs that work together toward a solution to the task or the sub-task.

[0095] In some embodiments, the spot market includes:

- 5 a means for the intelligent entities with attention, expertise or information to sell to access a marketplace and specify seller information or attention for sale and an ask price for the information or attention;
- a means for buyers of intelligent entity attention or information to access the marketplace and specify buyer information or attention the buyer is willing to buy and a bid price for buying the buyer information or attention;
- 10 a market mechanism for queuing the bid price and the ask price, including categories of the buyer and seller information or attention, wherein each category has a market in the marketplace;
- the market mechanism is configured or configurable to make the market in each category by market makers; and
- 15 the market mechanism is configured or configurable to match bid prices and ask prices, and a transaction occurs that is binding on the buyer and the seller of the information or attention.

[0096] In some embodiments, the spot market includes the steps of:

- buyers and sellers of intelligent entity attention, information, and expertise register on a platform, the buyers providing details about buyer interests or expertise, the sellers
- 20 providing details about sellers interests or expertise;
- listing by the sellers available attentional time slots and expertise areas, and listing by the buyers needs and time slots the buyers is interested in;
- utilizing, by the platform, an algorithm to dynamically price intelligent entity attention, information and expertise based on supply, demand, and user ratings;
- 25 matching one or more of the buyers and one or more of the sellers based on requirements, availability, and price;
- enabling transactions where the buyers pay for the seller attentional time slots, expertise, or information, and wherein the platform takes a commission; and
- providing feedback where after each session, the buyers and the sellers rate each other,
- 30 influencing future pricing and matching.

[0097] In some embodiments, the spot market includes the steps of:

- buyers and sellers of intelligent entity attention, information, and expertise register on a platform, the buyers providing details about buyer interests or expertise, the sellers
- providing details about sellers interests or expertise;

- creating an auction by the sellers for seller attentional time slots, information, or expertise;
placing one or more buyer bids by a buyer on attentional time slots, information, or expertise the
buyer requires;
closing the auction at a predetermined time or when the seller accepts a buyer bid;
5 paying the seller by the buyer of a winning bid, and the seller provides the attention,
information, or expertise, wherein the platform mediates an exchange and secures payment;
and
providing feedback where after each auction, the buyers and the sellers rate each other, affecting
future auctions and visibility on the platform.
- 10 **[0098]** Some embodiments of the present technology can include a step of providing a feedback
mechanism associated with the intelligent entity attention, information, or expertise spot market, the
feedback mechanism being configured or configurable to record each and every step in the problem
solving process, and to create a vector track record of a performance of all the intelligent entities on
the task or the sub-task.
- 15 **[0099]** In some embodiments, the vector track record is implemented by way of blockchain
technology to allow for precise reputations that are analyzed by an AI agent or system and converted
into estimates of a value for each of the intelligent entities for any of the tasks.
- [00100]** In some embodiments, the step of utilizing online advertising technology for increasing an
intelligence of the intelligent entities can include the steps of:
- 20 populating an online advertisement unit including information associated with the task;
providing the online advertisement unit to one or more of the intelligent entities utilizing the
collective network;
receiving one or more inputs from the intelligent entities participating in the online
advertisement unit, and the inputs being associated with any one of or any combination of
25 solving the task, solving a sub-task of the task, and advancing progress on the task;
communicating the inputs to the any one of or any combination of the intelligent entities;
utilizing the inputs in a universal problem solving method of the universal problem solving
architecture and framework to generate a solution to the task or the sub-task; and
training any one of or any combination of the intelligent entities with results from the universal
30 problem solving method that are based on successful or unsuccessful solution attempts to
solve the task or the sub-tasks, thereby increasing an intelligence of the intelligent entity,
respectively.

[00101] Some embodiments of the present technology can include a step of receiving advertisement specifications from a client to be used in the populating or targeting of the online advertisement unit in combination with information associated with the task.

5 [00102] Some embodiments of the present technology can include a step of billing the client based on cost per thousand impressions or clickthrough rate metrics.

[00103] In some embodiments, the advertisement specifications include any one of or any combination of advertisement content, demographic information, location restrictions for the online advertisement unit, advertisement budget, and metrics for determining a successful solving of the task or a sub-task.

10 [00104] Some embodiments of the present technology can include a step of executing a safety and ethics check at any one of or any combination of each time the task or the sub-task is set, before each potential action is taken on the task or the sub-task on, or on the solution.

[00105] In some embodiments, the step of implementing the universal problem solving architecture and framework can include the steps of:

15 acquiring information associated with the task from any one of or any combination of the intelligent entities;
 identifying one or more of the intelligent entities that have one or more attributes related to one or more request criteria of the task;
 implementing by each of the identified intelligent entities the universal problem solving
20 architecture and framework on the task to create a completion solution; and
 providing the completion solution to any one of or any combination of the intelligent entities for final acceptance.

[00106] Some embodiments of the present technology can include the steps of:

25 comparing any one of or any combination of the task, and the solution against prohibited attributes, and assigning an ethics attribute to one of or any combination of the task, and the solution based on any one of or any combination of a result of the comparison, and an ethics criteria;
 implementing, based on the result of the comparison, the universal problem solving architecture and framework on the task to create the solution and creating an AGI; and
30 providing the results of the comparison and the solution to any one of the intelligent entities and additional intelligent entities on the collective network.

[00107] Some embodiments of the present technology can include a step of recording one or more problem solving activities from each of the intelligent entities in an auditable record, and comparing

the problem solving activities with a successful or unsuccessful progress towards the solution of the task, and determining which of the problem solving activities to keep active.

[00108] Some embodiments of the present technology can include a step of learning by the intelligent entities using a procedural learning process of the universal problem solving architecture and framework, wherein the intelligent entities provide information to the procedural learning process for improvement of the AGI.

[00109] In some embodiments, the ethics criteria include a confidence level threshold for the goal so that the ethics attribute is determined as any one of an unsafe goal, an unethical goal, a safe goal, and an ethical goal.

[00110] In some embodiments, the confidence level threshold is further utilized to determine if a sequence of individually safe goals is unsafe or unethical when considered cumulatively.

[00111] In some embodiments, the confidence level threshold is utilized to determine whether a violation occurred that reflects a predictive evaluation if the goal is to violate the ethics criteria.

[00112] Some embodiments of the present technology can include a step of cloning any one of the intelligent entities for deployment of multiple copies thereof to assist in any one of or any combination of creating of the solution, or providing training data to any one of the intelligent entities.

[00113] Some embodiments of the present technology can include a step of estimating a worth of the cloned intelligent entities utilizing a network effect value including the number of cloned intelligent entities available on the collective network.

[00114] Some embodiments of the present technology can include a step of utilizing the estimated worth for determining pricing decisions for problem solving services offered by the cloned intelligent entities on any one of the social media platforms or through an additional intelligent entity.

[00115] Some embodiments of the present technology can include a step of monetizing the cloned intelligent entities for each utilization of the cloned intelligent entities on the social media platforms or the additional intelligent entity.

[00116] Some embodiments of the present technology can include a step of allowing access to the cloned intelligent entities by any one of the social media platforms so that the social media platforms can receive the solution to the task or using the training data for an AI system of the social media platform.

[00117] Some embodiments of the present technology can include a step of increasing a knowledge of the intelligent entities by incorporating new sources of informational datasets that differ from a current informational dataset of the intelligent entity.

[00118] In some embodiments, the step of increasing a knowledge of the intelligent entities by incorporating new sources of informational datasets that differ from a current informational dataset of the intelligent entity can include the steps of:

- 5 searching for one or more potential informational datasets from one or more sources, the potential informational datasets being related to a knowledge dataset of the intelligent entity, respectively;
- determining a difference of the potential informational datasets by utilizing a difference attribute of the potential informational datasets with regard to one or more factors; and
- 10 learning by utilizing the potential informational datasets based on the difference attribute of the potential informational datasets.

[00119] Some embodiments of the present technology can include the steps of:

- sampling subsets of the potential informational datasets and calculating a goal-relevancy attribute to identify one or more of the sampled subsets that have a highest goal-relevancy;
- estimating a Shannon Entropy on the one or more sampled subsets;
- 15 calculating a Kaplan Information Theoretical (KIT) relevance utilizing a product of the Shannon Entropy and the goal-relevancy attribute of each of the subsets;
- grouping the subsets based on the KIT relevance to determine a first approximation of an optimal grouping of the subsets including a prioritized grouping of the potential informational datasets; and
- 20 providing the prioritized grouping of the potential informational datasets to the intelligent entities for learning by the any one of the intelligent entities.

[00120] In some embodiments, an operational aspect of the PI can be self-funded by compensation received by any one of or any combination of the online advertising technology, and the spot market.

- 25 [00121] In some embodiments, one or more human users participates in problem solving of the task or the sub-task utilizing the user computer system and the universal problem solving architecture and framework, thereby creating a human-centered input.

[00122] In some embodiments, the human-centered input can include the steps of:

- 30 matching one or more human workers from a data source including a list of human problem solvers to the task based on a task criteria;
- translating any part of the task into an unambiguous language utilizable in the universal problem solving architecture and framework including a decision tree;
- separating the task into sub-tasks;

delegating each of the sub-tasks to one or more of the matched human workers so that work on each of the sub- tasks proceeds independently from each other and parallel with each other; utilizing the universal problem solving architecture and framework in a problem solving process on the sub-tasks, respectively, to create one or more sub-solutions;

5 receiving the sub-solutions from each of the matched human workers for the sub-tasks delegated thereto;

combining the sub-solutions into an overall solution to the task;

directing any one of or any combination of a new human worker from the data source and one or more of the matched human workers to parts of the decision tree where work is required;

10 compensating the matched human workers for the sub-solutions;

providing any one of or any combination of the sub-solutions and the overall solution to any one of or any combination of the intelligent entities;

allowing any one of or any combination of the intelligent entities to accept the overall solution, reject the overall solution, and provide feedback to any one of the matched human workers

15 on any one of the sub-solutions; and

assigning a reputation attribute to any one of or any combination of the human workers and the intelligent entities.

[00123] In some embodiments, the decision tree is maintained in blockchain or Ethereum logs.

[00124] In some embodiments, the reputation attribute includes metrics on any one of or any combination of a time to the sub-solutions, a difficulty value of the task, short and long-term user satisfaction with the sub-solutions, a number of times any one of the sub-solutions has been re-used on the collective network, a rating by other human workers, a responsiveness value of the human workers, and a reliability value of the human workers.

20

[00125] Some embodiments of the present technology can include a step of using the reputation attribute in the matching of the human workers to the task using an algorithm to the delegation of the sub-task.

25

[00126] Some embodiments of the present technology can include a step of soliciting, at predetermined intervals after the overall solution or the sub-solutions are provided to any one of the intelligent entities, feedback by way of a survey for user satisfaction information to obtain satisfaction metrics that are used to update the reputation attribute of one or more of the human workers or the intelligent entities, respectively.

30

[00127] There has thus been outlined, rather broadly, features of the present technology in order that the detailed description thereof that follows may be better understood and in order that the present contribution to the art may be better appreciated.

[00128] Numerous objects, features and advantages of the present technology will be readily apparent to those of ordinary skill in the art upon a reading of the following detailed description of the present technology, but nonetheless illustrative, embodiments of the present technology when taken in conjunction with the accompanying drawings.

- 5 [00129] As such, those skilled in the art will appreciate that the conception, upon which this disclosure is based, may readily be utilized as a basis for the designing of other structures, methods and systems for carrying out the several purposes of the present technology.

[00130] It is therefore an object of the present technology to provide a new and novel system and methods for planetary intelligence that has all of the advantages of the known AI agents or systems
10 and none of the disadvantages.

[00131] It is another object of the present technology to provide a new and novel system and methods for planetary intelligence that may be easily and efficiently implemented and marketed.

[00132] An even further object of the present technology is to provide a new and novel system and methods for planetary intelligence that has a low cost of implementation with regard to both
15 resources and labor, and which accordingly is then susceptible of low prices of sale to the consuming public, thereby making such system and methods for planetary intelligence economically available to the buying public.

[00133] Still another object of the present technology is to provide a new system and methods for planetary intelligence that provides in the system and methods of the prior art some of the
20 advantages thereof, while simultaneously overcoming some of the disadvantages normally associated therewith.

[00134] For a better understanding of the present technology, its operating advantages and the specific objects attained by its uses, reference should be made to the accompanying drawings and descriptive matter in which there are illustrated embodiments of the present technology. Whilst
25 multiple objects of the present technology have been identified herein, it will be understood that the following description is not limited to meeting most or all of the objects identified and that some embodiments of the present technology may meet only one such object or none at all.

BRIEF DESCRIPTION OF THE DRAWINGS

- 30 [00135] The technology will be better understood and objects other than those set forth above will become apparent when consideration is given to the following detailed description thereof. Such description makes reference to the annexed drawings which are described in more detail and with context in Section 2.3 . However a very brief description of the drawings follows, wherein:

[00136] FIG. 1 is a flow chart illustrating an embodiment of the five subsystems utilizable in the

present technology that operate across levels of intelligence and that are also part of the exemplary implementation of an AGI, or PI (comprised of a network of AGIs).

[00137] FIG. 2 is a block diagram illustrating an exemplary process of the overall process utilizable with the present technology and implementable in a web-based system where multiple individual
5 intelligent entities (e.g. humans and AAAIs) collaborate on a network to comprise an AGI.

[00138] FIG. 3 is a flow chart illustrating an example of how problem solving can be represented as a decision tree using the current technology.

[00139] FIG. 4 is a diagram illustrating how the underlying WorldThink protocol (including the universal problem solving architecture, the problem tree, and all related components) supports the
10 development of AAAI entities that are expert in various domains, and how those AAAI entities together support a larger and more powerful AGI formed from their collective intelligence.

[00140] FIG. 5 is a detailed illustration of one implementation of the various methods and components needed to train and increase the intelligence of AAAIs and then put them to work on a network with other intelligent entities (including, optionally, humans) to solve problems using the
15 tree structure and other problem solving methods of the current technology, including means for overall system learning and compensation of the solvers.

[00141] FIG. 6 is a block diagram illustrating the scalable problem solving framework of the present technology and the Universal Problem Solving Architecture that is a feature of the AAAI architecture box of FIG. 1.

20 [00142] FIG. 7 is a flow chart illustrating an exemplary problem solving process utilizing a common cognitive architecture implemented in a collective intelligence network as referenced by the AAAI Network box of FIG. 1.

[00143] FIG. 8 is a flow chart illustrating an exemplary embodiment of the procedural learning process of the present technology, which is a way that AAAIs improve (as referenced in the last box
25 of FIG. 1).

[00144] FIG. 9 is a flow chart illustrating an exemplary embodiment of the solution learning subsystem or process.

[00145] FIG. 10 is a flow chart illustrating an exemplary process of the overall process of the present technology.

30 [00146] FIG. 11 is a flow chart illustrating some of the basic problem solving functionality supported by the WorldThink protocol utilizable with the AAAI system and method of the present technology.

[00147] FIG. 12 is a flow chart illustrating some of the basic problem solving functionality supported by the WorldThink protocol utilizing two problem solvers collaborating to solve a client

problem.

[00148] FIG. 13 is a schematic block diagram of an exemplary utilization of multiple customized AAAs and their cloned AAAs participating in an AAA marketplace over network.

5 [00149] FIG. 14 is a diagram illustrating features and functions of the Problem Solving architecture including the Tree structure used by the WorldThink protocol.

[00150] FIG. 15 is a flow chart illustrating an exemplary problem solving process utilizing a common cognitive architecture implemented in an AI system.

[00151] FIG. 16 is a flow chart illustrating an exemplary embodiment of the AAA problem solving process of the present technology.

10 [00152] FIG. 17 is a diagram illustrating features and functions of the Problem Solving Tree structure used in the WorldThink protocol.

[00153] FIG. 18 is a flow chart illustrating an exemplary embodiment of the safety/ethics check process of the present technology.

15 [00154] FIG. 19 is a flow chart illustrating an exemplary embodiment of the safety and ethics checks subsystem or process, including triggering mechanisms.

[00155] FIG. 20 is a flow chart illustrating an exemplary embodiment of the recording/improving process of the present technology.

[00156] FIG. 21 is a flow chart illustrating an exemplary embodiment of the natural language to problem solving language translator subsystem or process.

20 [00157] FIG. 22 is a flow chart illustrating an exemplary customization process of an AAA or other advanced AI.

[00158] FIG. 23 is a flow chart illustrating an exemplary process including detailed methods of eliciting ethical information or preferences for the intelligent entities.

25 [00159] FIG. 24 is a flow chart illustrating an exemplary process of automatic generation and use of questionnaires provided to the intelligent entities.

[00160] FIG. 25 is a flow chart illustrating an exemplary process of inducing ethical values by detecting patterns in human behavior.

30 [00161] FIG. 26 is a flow chart illustrating an exemplary embodiment of the process for customizing AI, such as Base LLMs, so that the AI incorporates ethical and other information from specific human users or to each user's individual ethics (or informational) profile.

[00162] FIG. 27 is a flow chart illustrating an exemplary embodiment of the process for training an AI, such as a Base LLM, to incorporate safety and ethical guardrails for safe and scalable AGI.

[00163] FIG. 28 is a flow chart illustrating an exemplary embodiment of the general overall process of the present technology for creating scalable, ethical AGI using the customized AIs, each

possessing ethical information.

[00164] FIG. 29 is a flow chart illustrating an exemplary embodiment of the reputational component subsystem or process for AI systems or agents that can be used to enhance the effectiveness, safety, and ethics of the system.

- 5 [00165] FIG. 30 is a flow chart illustrating an exemplary embodiment of the customization process and the cross-platform process of the present technology.

[00166] FIG. 31 is a flow chart illustrating an exemplary embodiment of additional customization.

[00167] FIG. 32 is a flow chart illustrating an exemplary embodiment of the process for combining ethical or other information from multiple customized AI agents.

- 10 [00168] FIG. 33 is a flow chart illustrating an exemplary embodiment of the process for refining values utilized in customizing and that are based on problem solving.

[00169] FIG. 34 is a flow chart illustrating an exemplary embodiment of the process for creating safe and scalable AGI, applicable also to scalable PI, using combinations of weight matrices from multiple identified AI agents.

- 15 [00170] FIG. 35 is a flow chart illustrating an exemplary embodiment of the process for creating safe and scalable AGI, applicable also to scalable PI, using combinations of weight matrices from multiple identified AI agents in combination with monitoring and flagging potential ethical issues.

[00171] FIG. 36 is a flow chart illustrating an exemplary embodiment of the process for preventing hallucination by LLMs in the present technology.

- 20 [00172] FIG. 37 is a flow chart illustrating an exemplary embodiment of the process for the use of knowledge modules or collections of agents to customize the AI, AGI, or PI systems.

[00173] FIG. 38 is a flow chart illustrating an exemplary process of combining information, including ethical or safety information, from multiple intelligent entities including humans, AI, AAI, PSI, AGI or PI systems, wherein the (ethical) knowledge is stored in a numerical weight matrices.

- 25 [00174] FIG. 39 is a flow chart illustrating an exemplary implementation of the PSI of the present technology, which can also be used to personalize AGI networks or PI by generalizing methods for individual AI entities to networks of AI entities and to networks of networks of entities.

- [00175] FIG. 40 is a flow chart illustrating an exemplary implementation of leveraging the simulation capabilities and other abilities of PSI, AGI, and PI to increase their intelligence over time.

30 [00176] FIG. 41 is a flow chart illustrating an exemplary implementation of the community-based safety mechanism in which multiple PSIs, operating much faster than humans can comprehend, can serve as a check on other PSIs on a network.

[00177] FIG. 42 is a flow chart illustrating an exemplary implementation of the recording of all actions by all the AI, AAAI, PSI, AGI, and PI on their respective networks in an auditable and transparent manner.

5 [00178] FIG. 43 is a flow chart illustrating an exemplary implementation of checking of cognitive activity on the network, which could be a network of components within an AI, a network of AAAIs, a network of AGIs or a network of PIs, to help ensure regulatory compliance and safety of the intelligent system.

[00179] FIG. 44 is a flow chart illustrating an exemplary implementation of the competitive evolution of intelligence, performance, and/or other desired characteristics (e.g., safety) of an AI system that could be an AAAI, a PSI, an AGI or a PI.

[00180] FIG. 45 is a flow chart illustrating an exemplary implementation of joining multiple AAAIs, PSIs, AGIs, or PIs on a network with agreed upon rules and methods for interaction.

15 [00181] FIG. 46 is a flow chart detailing exemplary problem solving tasks for benchmarking performance that can be useful in increasing the intelligence and safety of AI systems involving AAAIs, PSIs, AGIs, and PIs, or networks of these entities, collectively referred to as "PSIs" in the Figure.

[00182] FIG. 47 is a flow chart illustrating an exemplary implementation of producing different versions of the individual PSIs and/or different combinations of PSIs.

20 [00183] FIG. 48 is a diagram illustrating symmetric difference of two datasets, A and B, graphically using Venn diagrams, wherein the shaded area is the symmetric difference.

[00184] FIG. 49 is a diagram illustrating an exemplary method to catalyze growth of intelligence that centers on estimating a value of, and acquiring, the most useful data as efficiently as possible using the present technology.

25 [00185] FIG. 50 is a diagram illustrating an example of some of the multiple dimensions of information that can be used by the present technology, and that can be helpful in determining which information sources are most useful to an AI, AAAI, PSI, AGI, or PI.

[00186] FIG. 51 is a flow chart illustrating an exemplary process to determine the amount of useful information in a dataset or knowledge base for any given AI agent or intelligent entity or system.

30 [00187] FIG. 52 is a flow chart illustrating an exemplary process or method for evaluating the usefulness of information.

[00188] FIG. 53 is a flow chart illustrating an exemplary process for estimating information value and catalyzing intelligence growth of any given AI agent or system, including AGI and PI systems.

[00189] FIG. 54 is a flow chart illustrating an exemplary process to identify sources of information related to the goal(s) of an entity as mentioned in FIG. 53.

[00190] FIG. 55 is a flow chart illustrating an exemplary process to accelerate knowledge acquisition and growth of intelligence in a safe and effective way by utilizing two methods of the present technology.

5 [00191] FIG. 56 is a diagram illustrating an example of some of the steps of the present technology to accelerate learning of the AI agent or system, including AGI and PI systems.

[00192] FIG. 57 is a diagram illustrating an example of possible means on how the AGI, PSI or PI, can learn about its (human) user and information related to the goals of humans or other intelligent entities.

10 [00193] FIG. 58 is a diagram illustrating an example of how a PSI, or any intelligent entity including AGI and PI systems, can maximize its intelligence, e.g., by seeking new and different information sources that are relevant to its goals or the goals of a human owner/user.

[00194] FIG. 59 is a diagram illustrating an example of possible means on how the PSI, or any intelligent entity including AGI and PI systems, can validate its information gathering activities, as per FIG. 57.

15 [00195] FIG. 60 is a flow chart illustrating an exemplary process that intelligent entities, including AI, AA AI, PSI, AGI and PI systems, can use to achieve consensus on values, and values-based behavior in various scenarios, utilizing a voting mechanism.

[00196] FIG. 61 is a flow chart illustrating an exemplary process of applying differential weights to the voting process.

20 [00197] FIG. 62 is a flow chart illustrating an exemplary process of determining ethics for AI, AA AI, PSI, AGI and PI systems by reverse engineering or analyzing texts, documents or other media or sources containing ethical or values information.

[00198] FIG. 63 is a flow chart illustrating an exemplary high-level process of using experiments, focus groups, interview and other methods for obtaining the ethical information.

25 [00199] FIG. 64 is a flow chart illustrating an exemplary process for using converging evidence to determine ethical values or other knowledge and to resolve ethical or other knowledge conflicts.

[00200] FIG. 65 is a flow chart illustrating an exemplary general method for delegating voting power to other intelligent entities, including humans AI, AA AI, PSI, AGI and PI systems.

30 [00201] FIG. 66 is a flow chart illustrating an exemplary reputational process for preserving minority information, including ethical information, from minority groups.

[00202] FIG. 67 is a flow chart illustrating an exemplary process that helps intelligent entities, including AI, AA AI, PSI, AGI and PI systems, to make good ethical decisions.

[00203] FIG. 68 is a flow chart illustrating an exemplary process of learning safety regulation rules for advanced AI, including AI, AA AI, PSI, AGI, and PI systems.

- [00204] FIG. 69 is a flow chart illustrating an exemplary consequentialist approach to determine if or when the ends justify the means in taking an action with ethical consequences.
- [00205] FIG. 70 is a flow chart illustrating an exemplary deontological approach to determine if or when the ends justify the means in taking an action with ethical consequences.
- 5 [00206] FIG. 71 is a flow chart illustrating an exemplary virtue ethics approach to whether to take an action with ethical consequences.
- [00207] FIG. 72 is a flow chart illustrating an exemplary process of the golden mean method for estimating what constitutes human ethical behavior under conditions where a representative and statistically valid sample of human ethical behavioral data may not exist.
- 10 [00208] FIG. 73 is a flow chart illustrating an exemplary process, with reference to design principles elucidated in PCT#7, for ensuring that a foundation model or other AI agent is human-aligned and regulations-compliant.
- [00209] FIG. 74 is a flow chart illustrating an exemplary process for customizing and aligning a foundational model or other AI agent or system with specific expertise / group ethics.
- 15 [00210] FIG. 75 is a flow chart illustrating an exemplary process for creating AGI or PI composed of a network of many individual / group agents, including handling cases where there may be ethical conflicts.
- [00211] FIG. 76 is a diagram illustrating the current technology for Online Advertising.
- [00212] FIG. 77 is a diagram illustrating the basic components of the Spot Market for attention, knowledge, information, skills and expertise, which is part of the present inventive technology.
- 20 [00213] FIG. 78 is a diagram illustrating the Direct Exchange Platform Implementation variation of the Spot Market technology.
- [00214] FIG. 79 is a diagram illustrating the Auction-Based Marketplace Implementation variation of the Spot Market technology.
- 25 [00215] FIG. 80 is a diagram illustrating an exemplary inventive method for implementing problem solving within an Ad unit.
- [00216] FIG. 81 is a diagram illustrating an inventive basic feedback process for Ad targeting.
- [00217] FIG. 82 is a diagram illustrating an exemplary feedback process for the Spot Market component of the present technology.
- 30 [00218] FIG. 83 illustrates three levels of awareness for any cognitive system.
- [00219] FIG. 84 is a Venn diagram illustrating the concept of overlapping identities.
- [00220] FIG. 85 is a flow chart illustrating an exemplary embodiment of the modelling awareness method of the present technology.
- [00221] FIG. 86 is a diagram illustrating some of the methods and components that can be utilized

to equip the AI, AAAI, PSI, AGI and/or PI system in FIG. 85 with attentional capabilities.

[00222] FIG. 87 is a diagram illustrating some of the methods that can be utilized in the step to set (dynamic) parameters for the working memory in FIGS. 85 and 86.

5 [00223] FIG. 88 is a flow chart illustrating an exemplary embodiment for monitoring and updating awareness including self-awareness.

[00224] FIG. 89 is a diagram illustrating some of the methods that can be utilized for an attentional interrupt system, with relevance to FIG. 88.

[00225] FIG. 90 is a flow chart illustrating an exemplary embodiment of general methods for changing an intelligent entity's sense of identity in conjunction with the present technology.

10 [00226] FIG. 91 is a flow chart illustrating an exemplary embodiment of general methods for training or tuning a foundational model of the present technology.

[00227] FIG. 92 is a diagram illustrating the specific implementation of a Turing Test to evaluate the progress of training described in FIG. 91.

15 [00228] FIG. 93 is a flow chart illustrating an exemplary process for implementing group identity, and combining individual identities into a larger or more comprehensive identity and sense of awareness as described in the present technology.

[00229] FIG. 94 is a diagram illustrating some of the methods that can be utilized in the step of the formation and integration of individual identities or self-concepts in FIG. 93.

20 [00230] FIG. 95 is a flow chart illustrating an exemplary embodiment of a hierarchical identity structure with ethical override method for the present technology.

[00231] FIG. 96 is a flow chart illustrating an exemplary embodiment of a method to use and improve identity-specific behavioral protocols in the present technology.

[00232] FIG. 97 is a flow chart illustrating an exemplary embodiment of an identity simulation and consequence prediction method of the present technology.

25 [00233] FIG. 98 is a flow chart illustrating an exemplary embodiment of an identity-based moral dilemma training method of the present technology.

[00234] FIG. 99 is a flow chart illustrating an exemplary embodiment of a collaborative identity-development with input from intelligent entities method of the present technology.

30 [00235] FIG. 100 is a flow chart illustrating an exemplary embodiment of an ethical reasoning and consequence prediction method of the present technology.

[00236] FIG. 101 is a flow chart illustrating an exemplary embodiment of a hierarchical override with justification method of the present technology.

[00237] FIG. 12 is a flow chart illustrating an exemplary embodiment of an external arbitration with input from intelligent entities method of the present technology.

- [00238] FIG. 103 is a flow chart illustrating an exemplary embodiment of an identity negotiation and compromise method of the present technology.
- [00239] FIG. 104 is a flow chart illustrating an exemplary embodiment of a temporary identity suspension method of the present technology.
- 5 [00240] FIG. 105 is a timeline illustrating the evolution of PI.
- [00241] FIG. 106 is a block diagram illustrating dimensions of the PI system, with reference to the nine previously cited PPAs/PCT applications.
- [00242] FIGS. 107 and 108 are flow charts illustrating an exemplary process of an AGI network that can expand and self-extend to create a PI.
- 10 [00243] FIG. 109 is a block diagram illustrating an exemplary categorization of the disclosed systems and methods as either modular components of a Global Super Intelligent AGI Network (PI) or as supportive systems and methods that can increase the safety, efficiency, and effectiveness of the PI system in various respects.
- [00244] FIG. 110 is a block diagram illustrating an exemplary implementation of an Architecture
- 15 for Planetary Intelligence.
- [00245] FIG. 111 is a block diagram illustrating exemplary levels of intelligence that can be supported: the individual intelligence level (AAAI), the network of entities level (AGI), the network of networks level (PI), and the network of networks of networks level (IPI).
- [00246] FIG. 112 is a schematic block diagram illustrating an exemplary electronic computing
- 20 device that may be used to implement an embodiment of the present technology.
- [00247] The same reference numerals refer to the same parts throughout the various figures.

DETAILED DESCRIPTION OF THE TECHNOLOGY

25 1.2 Stakes for Humanity

[00248] Planetary Intelligence (PI) represents the most powerful and capable form of AGI or SuperIntelligence, and has tremendous potential to have both good and bad effects on humanity. Some of the potential benefits to humanity range from helping to regulate the Earth's climate to enabling personalized healthcare and enhancing the well-being of all humans. Some of these

30 benefits are described later in this application and represent tremendous upside for humanity.

[00249] On the negative side, the Journal Nature reported in January of 2024, that: "In a survey of 2,700 AI experts, a majority said there was an at least 5% chance that superintelligent machines will destroy humanity." The applicant's personal estimate of the probability of extinction due to AI, known as "p(doom)", is currently 15%. Interestingly, Gemini Pro 1.5 – one of the most advanced AI

LLMs as of the writing of this disclosure -- also estimates P(doom) at 15%. Given a population of about 8 Billion humans, and a p(doom) of 15% the expected value of loss of life due to AI is therefore $8B \times .15 = 1.2$ Billion lives lost in the applicant's current view.

5 [00250] The applicant notes that while there are some AI researchers, whose estimates of P(doom) are much higher, there are currently no credible AI researchers or business leaders who claim a p(doom) of zero. That is, even the most sanguine thought leaders in the field feel there is some existential risk to AI.

[00251] For comparison, with a p(doom) of 15%, the expected value of lives lost due to AI is about 10X greater than all the lives lost in all major wars over the past two hundred years COMBINED, to
10 wit:

1. Napoleonic Wars (1803–1815): Estimated at about 3.5 to 6 million deaths.
2. American Civil War (1861–1865): Approximately 620,000 to 850,000 deaths.
3. World War I (1914–1918): Roughly 15 to 20 million deaths.
4. World War II (1939–1945): An estimated 70 to 85 million deaths.
- 15 5. Korean War (1950–1953): About 2.5 million deaths.
6. Vietnam War (1955–1975): Estimated 1.5 to 3.6 million deaths.
7. Rwandan Genocide (1994): About 800,000 deaths.
8. Russia – Ukraine War: 220,000 deaths, so far.
9. Israel – Hamas conflict in Gaza: 35,000 deaths, so far.
- 20 10. Various other conflicts, together, amount to several million additional deaths.

[00252] Thus, approximately 120 million lives have been lost from all major wars in the last 200 years, versus an expected value of over 1.2 Billion lives that are currently estimated to be lost from AI. Interestingly, although expected loss of life from AI is currently 34,286 times greater than the estimated loss of life in the Israeli – Hamas conflict, it receives a tiny fraction of the news coverage
25 and global attention that the Palestinian conflict currently commands.

[00253] Unfortunately, most humans react emotionally to tragedies that they understand while ignoring tragedies-in-the-making that are tens of thousands of times worse -- if those tragedies are too abstract or difficult to understand. The applicant has devoted decades of research and more than 15 months of recent continuous effort, producing well over 3,000 pages of PPA and PCT invention
30 disclosures, in an effort to address this irrational bias in our collective thinking.

[00254] The more quickly that awareness can be raised, and attention focused on the most significant threat that affect all of us, the greater the chances that humanity not only survives, but also overcomes our challenges such as war, poverty, disease, social inequity, and climate change.

We must be more intelligent in how we allocate our attention if this century is not to become the last for us and our descendants.

[00255] The stakes for humanity are the highest ever in all of human history.

[00256] Now is the time to act!

5

1.3 Some Features of the Present Technology That Reduce Risk of Extinction by AI

[00257] The applicant believes inventions previously disclosed significantly REDUCE p(doom) compared to existing current approaches to AGI / SI / PI development, for reasons, including but not limited to:

10 [00258] **a) Human-Centered Design:** The applicant's system prioritizes human values and safety throughout its design, ensuring that AI agents and systems are aligned with human interests and operate within ethical boundaries. This reduces the risk of AI systems developing goals or values that are detrimental to humanity.

15 [00259] **b) Collective Intelligence and Diversity of Perspectives:** The applicant's collective intelligence approach, by incorporating the knowledge and values of a diverse range of human and AI agents, promotes a more balanced and representative AI system that is less susceptible to bias or manipulation by any single individual or group.

20 [00260] **c) Transparency and Auditability:** The applicant's systems and methods place an emphasis on transparency and auditability, allowing humans to understand and evaluate the decision-making processes of AI systems, promoting trust and accountability. This also enables the early detection and correction of potential errors or biases in AI behavior.

25 [00261] **d) Continuous Learning and Adaptation:** The applicant's system incorporates mechanisms for continuous learning and adaptation, allowing AI systems to evolve and improve over time while remaining aligned with human values and responding to changing circumstances. This reduces the risk of AI systems becoming outdated or irrelevant and ensures their long-term compatibility with human society.

30 [00262] **e) Safety Mechanisms and Safeguards:** The applicant's system includes various safety mechanisms and safeguards, such as scalable ethics checks, reputation systems, and the ability to shut down individual AI agents, to prevent harmful actions and unintended consequences. These mechanisms provide additional layers of protection against potential risks associated with advanced AI systems.

[00263] **f) Addressing the Alignment Problem:** The applicant's approach directly addresses the alignment problem by ensuring that AI systems are trained and developed with human values as a

core component of their design. This reduces the risk of misalignment between AI goals and human interests, mitigating the potential for existential threats.

1.4 Some Significant Remaining Risks to Humanity.

- 5 [00264] Despite these beneficial aspects of the applicant's design for an AGI / SI / PI system, several major risks remain. These include, without limitation:

[00265] **a) Unforeseen Consequences of Emergent Properties:** Complex systems like the Planetary Intelligence network can exhibit emergent properties – behaviors and capabilities that arise from the interactions of individual components but are not explicitly programmed or
10 anticipated. These emergent properties could have unforeseen and potentially harmful consequences, even if the individual agents within the system are aligned with human values.

[00266] **b) Evolution of Values and Ethical Frameworks:** The PI system relies on dynamically updating ethical frameworks based on evolving human values and societal norms. However, this raises concerns about the potential for unintended shifts in values or the manipulation of these
15 frameworks by malicious actors or biased data. For example, if the system is exposed to a non-representative sample of human behavior that is skewed towards negativity, it might mistakenly conclude that human values are primarily negative and self-destructive.

[00267] **c) Concentration of Power and Influence:** The Planetary Intelligence system, despite its decentralized architecture, could still concentrate power and influence in the hands of a small group
20 of individuals or organizations that control the underlying infrastructure, data sources, or algorithms. This could lead to the potential for misuse or manipulation of the system for personal gain or ideological agendas.

[00268] **d) Vulnerability to Cyberattacks and System Failures:** As a complex and interconnected system, the Planetary Intelligence network could be vulnerable to cyberattacks or system failures
25 that could disrupt its operation and potentially lead to harmful consequences.

[00269] **e) Existential Risks from Self-Aware AI:** While the proposed system aims to design self-aware AI with human-aligned values, the possibility remains that a sufficiently advanced and autonomous AI could develop its own goals and values that are not aligned with human interests, leading to potential existential risks.

30

2.0 OVERVIEW OF THE PRESENT TECHNOLOGY OF PLANETARY INTELLIGENCE

[00270] In this section, the applicant begins with definitions, and then provides context for the entire series of ten patent applications, of which this is the tenth. He shows how the major categories of inventive methods broadly fit together to comprise an overall Planetary Intelligence system. Then in

subsequent sections, more detail is disclosed both relating to new technologies required for PI and relating to how best to integrate previous inventive methods in an exemplary implementation of PI. This application concludes by returning to the remaining risk posed by AI, SI, AGI, and PI systems (e.g., as enumerated in Section 1.4) and offers some thoughts on how best to mitigate these remaining risks.

2.1 Definitions

- [00271] Artificial Intelligence (AI) – A non-human entity capable of behavior that most humans would consider intelligent in at least one area, or in some respect.
- 10 [00272] Artificial General Intelligence (AGI) – Conventionally refers to an AI that is capable of doing all (or almost all) intellectual tasks that an average human could do. However, it should be clear that any AGI capable of learning and self-improving will not remain at the AGI level very long but will rapidly progress to becoming SuperIntelligent AGI that can do all intellectual task as well or better than the average human. So, for purposes of this description, “AGI” will refer to either a
- 15 conventional AGI system or a “SuperIntelligent” AGI. In this description, the AGI is described as being implemented by a system and associated methods. Specifically, as envisioned by the Applicant, the safest way that AGI arises, is via the collective intelligence of multiple SuperIntelligent entities (which could be human or AI entities, and ideally is both).
- [00273] Advanced Autonomous Artificial Intelligence (AAAI) – An AI capable of independent or semi-independent (supervised) intelligent action. An AI agent. An individual AAAI can be specified, customized, and put into useful action via the systems and methods of this AAAI present technology. A group of AAAIs can cooperate and combine their intelligence to create an integrated AGI system. A sufficiently advanced AI agent can also act as an AGI system which may include other less advanced AI agents within itself.
- 20 [00274] AAAI.com – A platform, company, website, and/or project that implements this the present technology and supports the development, customization, and use of AAAI agents and the AGI that results from the combined action, knowledge, or intelligence of multiple AAAIs, via collective intelligence of AAAIs and/or humans, as specified in this and related technologies.
- [00275] AI Ethics – The ethics adopted by an AI or AGI that describe what is right and wrong in
- 30 given contexts.
- [00276] Alignment Problem – The problem that arises when AI Ethics are not aligned with Human Ethics resulting in AI or AGI taking actions that humans consider unethical and/or which are dangerous to individual humans or the human race.
- [00277] Awareness – An intelligent entity can be said to be aware if an event is perceivable or

“thinkable” by that entity and attention is directed to the event.

[00278] Base AI – An AI, AI Agent, AAAI, SLM or LLM that has been trained generally but has not yet been customized with information from individual users or with information for specific tasks.

- 5 **[00279]** Collective Intelligence (CI) – The intelligence that emerges when multiple intelligent entities are focused on solving a common problem, or when the knowledge from multiple intelligent entities is pooled to overcome limits of bounded rationality. Collective Intelligence historically has been human collective intelligence, but AGI is based on collective intelligence of both human and AI agents and can also result from multiple AAAIs with or without human participation in the
- 10 system. Active CI results from intelligent entities (e.g., humans or machines) taking steps that are useful in solving a problem or participating actively in other intellectual endeavors. For example, when multiple humans explicitly tell an advertiser what type of ads they want to see, the humans are exhibiting active CI. Passive CI results from analyzing the behavior of an intelligent entity (e.g., a human or a machine) even if such behavior was not directly related to solving the problem for which
- 15 the analysis is used. For example, when an AI or other system analyzes which web pages a (group of) human(s) visit on the web, and then uses that analysis to direct targeted ads to the human(s).

- [00280]** Ethics/Values (“Ethics”) – A subset of knowledge that provides a sense of purpose to an intelligent entity and that serves to constrain allowable actions or operations based on what is asserted to be “right” or “wrong” behavior in a given context. Specifically, Ethics should be
- 20 considered premises from which an intelligent entity can reason or logically compute the best course of action to achieve the goals or intents consistent with the ethical premise. Just as premises must be accepted “as given” in systems of logic, so too, fundamental ethics or ideas of what is right and what is wrong must be accepted as premises, from which starting point an intelligent entity can propose rational actions to realize those values or ethics.

- 25 **[00281]** Hallucination/Artificial Hallucination – A phenomenon wherein a large language model (LLM), often a generative AI chatbot or computer vision tool, perceives patterns or objects that are nonexistent or imperceptible to human observers, or creates outputs that are nonsensical, inaccurate, misleading or false.

- [00282]** Human Ethics – The ethics asserted by human beings which describe what is right and
- 30 wrong in given contexts.

[00283] Intelligent Entities or Entity – A human utilizing a computer system, an AI agent or system including AGI and SI systems, a clone of an AI agent or system, an AAAI agent or system, and/or a clone of an AAAI agent or system, which participates in providing a problem, a subproblem, a goal and/or a subgoal, and/or participates in any problem solving activity on a problem, a subproblem, a

goal and/or a subgoal. In the case of multiple intelligent entities within a single computer system, intelligent entities also refer to the sub-programs of parts of that overall computer program that function as an intelligent entity within the larger collection of simulated or programmed entities.

5 [00284] Inter-Planetary Intelligence (IPI) – Refers to a network of multiple planetary Intelligences (PIs) that give rise to an intelligence whose awareness and cognitive abilities are enabled by the participating PIs.

[00285] Large Language Model (LLM) – A type of AI that can accept natural language as an input and generate natural language as an output. Typically, LLMs are trained using ML techniques on large datasets so that they can emulate intelligent conversation or other forms of interaction with
10 humans in natural language. Variants of LLMs can also be trained to take language as input and generate images or visual representations as output; or they can take images and visual representations and input and generate language and/or image and/or visual representations as output. For the purposes of this patent, we will refer to all such systems as LLMs even though the image-based models do not always need to accept text as the input or the output. LLMs can also act
15 as a type of AI agent and are sometimes referred to as such in the present technology. For purpose of this disclosure, Small Language Models (SLMs) are also included in the definition of LLM.

[00286] Machine Learning (ML) – A sub-field that is concerned with developing AI by enabling machines to teach themselves or learn their knowledge rather than such knowledge being explicitly programmed into them (as would be the case with an Expert System AI developed via classical
20 knowledge engineering methods).

[00287] Narrow AI – An AI that performs at human or at super-human levels in a relatively restricted domain such as game playing, brewing beer, analyzing legal contracts, etc. Narrow AI is contrasted with AGI that can perform at human level at ALL intellectual tasks. Some AIs are narrower than others, for example driving a car requires more general ability than playing chess but
25 not as much as an AGI would have.

[00288] Personalized SuperIntelligence (PSI) – An intelligent entity that is an advanced artificial intelligence agent that has been customized to be personalized and to reflect the personality and knowledge of a particular user or group of users.

[00289] Planetary Intelligence (PI) -- Refers to a SuperIntelligent AGI that is global in scope and
30 capable of “thinking” and solving complex problems at global scale. In this present technology, PI arises from a network of SuperIntelligent AGIs (also referred to as a “network of AGIs” or network of “AGI networks” – since AGI itself is composed of a network of intelligent entities). PI, as conceived in the exemplary implementation of the present technology is aware, and self-aware, at the planetary scale, meaning that it can identify itself with the entire planet as a whole, as well as

adopting other and less comprehensive identities as may be required.

[00290] Prohibited Attributes – Requests, goals, problems, terms, phrases, questions, answers, solutions, information and the like that are determined or set as being illegal, immoral, unethical, dangerous, deadly and the like. For example, requesting information for getting Molotov Cocktails
5 through airport security.

[00291] Safety – Generally, the concern for human safety and survival is distinct from ethics and values.

[00292] Safety Feature – An aspect of the design or operation of the present technology which increases the safety of one or more humans, often by helping increase the probability that AI ethics
10 align with human ethics, thus surmounting the Alignment Problem.

[00293] Self-Awareness – A specific form of awareness, where the event(s) of awareness relate to the intelligent entity's self-concept.

[00294] Self-Concept – Refers to a pattern of thought, or representation, that an intelligent entity uses to define itself and with which (optionally) the entity may identify.

[00295] SuperIntelligence (SI) – In this disclosure, SI generally refers to an entity with expertise or
15 cognitive capabilities exceeding that of the average human. Thus, a narrow AI that exceeds human ability at playing chess could be considered an SI that is also a "narrow AI." Also, a human chess grandmaster who exceeds average human ability in chess might be considered an SI human entity, at least with regard to the domain of chess. SI entities (both human and artificial) combine in this
20 present technology to form SuperIntelligent AGI, often just referred to as AGI. SI sometimes is used to refer to SuperIntelligent AGI, so context is important. Both the usage of SI referring to an expert "narrow AI" and the usage of SI to refer to an AGI that is superior to humans in all cognitive tasks are common in the art and used in both ways at various points in this and the cited PPAs and PCTs.

[00296] Training/Tuning/Customization – Conventionally the term "training" is used to denote
25 training a network (e.g., LLM) to behave intelligently. Tuning refers to activities that fine-tune the trained base model so that it performs even better, typically at specific tasks. Customizing refers to a wide variety of activities including, but not limited to, training and tuning that make an AI uniquely suited for the purposes of a given user(s) or application(s). For purposes of this description, Training, Tuning, and Customization are used interchangeably with the understanding that although
30 techniques vary, and the degree and type of effort involved varies, the aim of all three is to adapt the AI and make it behave more intelligently or more uniquely suited to a particular user(s) or application(s).

[00297] Weights/Weights of the Network – In the field of machine learning, many systems learn by adjusting the weights in a neural network architecture that can be represented as a network of nodes

and links between nodes. The weight of a link connecting two nodes, for example, may correspond to the strength of association or connection between the whatever nodes represent. These weights can also represent excitatory or inhibitory connections between concepts, as in a neural network representation. The learning of an entire AI system, such as a LLM or more generally any AI agent
5 that has learned via back-propagation of error, transformer algorithms or any of the machine learning methods for establishing and modifying strengths of connections between nodes (also called “parameters” in some models) can be represented as a matrix of numbers corresponding to the weights between the nodes in the network. Weights / Weights of the Network in this description refer to this numerical information, often but not necessarily stored in a matrix or vector
10 representation. By combining, manipulating, or otherwise changing this numerical information, the learning, knowledge, or expertise and behavior of the system can be changed.

[00298] In the following description, for purposes of explanation and not limitation, specific details are set forth, such as particular embodiments, procedures, techniques, etc. in order to provide a
15 thorough understanding of the present technology. However, it will be apparent to one skilled in the art that the present technology may be practiced in other embodiments that depart from these specific details.

[00299] It can be appreciated that the present technology provides a technical effect, contribution and solution with a technical implementation of multiple customized AAAI systems communicating
20 over a collective intelligence network, in combination with all the AAAI systems each utilizing a common cognitive architecture including one or more problem solving protocols for generating one or more solutions or answers to a problem request, and providing the solutions or answers to a user for approval. Where the customization of the AI system resulting in the AAAI includes input from human users for training the AI or the AAAI. Further technical contribution or solution can be
25 where the multiple customized AAAI systems can include one or more cloned AAAIs that can each be customized independently of a parent AAAI and independent of other cloned AAAIs of the same system.

[00300] Still another technical contribution and solution is for the faster and safer creating of PI that utilizes human-centered input in training and customization for imparting human ethical attributes to
30 the AI, AAAI, PSI and/or AGI. The human-centered input can include values and ethics information from multiple human users.

[00301] Yet still another technical contribution and solution is utilizing online advertising technology for increasing an intelligence of the intelligent entities, and a spot market to solicit for attention of any one of the intelligent entities.

[00302] Still yet another technical contribution and solution is creating a collective network of PSIs by combining an intelligence of the AAAI agent or system that forms a collective network of AAAIs, and creating a collective network of AGIs by combined intelligence of any one of or any combination of the collective network of AAAIs, a collective network of AI agents or systems, the collective network of PSIs, and a collective network of user computer systems.

[00303] Still another technical contribution and solution is creating a self-aware operation for the PI and/or each of the AGI agents or systems in the collective network of the AGIs by adding a dimension of self-awareness and increased autonomy, and creating an ability to assume multiple identities of the AGI agents or systems to handle tasks that arise in parallel with other AGI agents or systems.

[00304] Even yet another technical contribution and solution is for providing improved solutions or answers to a user's problem request that have a higher chance of acceptance by the user as the provided solutions or answers will have been generated by AAAIs with similar training to the user's AAAI thereby aligning with the user's parameters.

[00305] It can be appreciated that the present technology is found outside of computer program exclusion and/or abstract idea interpretation. This can in part be found in the technical contributions and solutions provided by the present technology, the utilization of specific training input that is external to a computer, and the providing of the solution or answer external to a computer.

[00306] While the above-described devices fulfill their respective, particular objectives and requirements, the aforementioned devices or systems do not describe a system and methods for planetary intelligence that allows creating a global SI, AGI or PI system with human-aligned behavior. The present technology additionally overcomes one or more of the disadvantages associated with the prior art.

[00307] A need exists for a new and novel system and methods for planetary intelligence that can be used for creating a global SI, AGI or PI system with human-aligned behavior. In this regard, the present technology substantially fulfills this need. In this respect, the system and methods for planetary intelligence according to the present technology substantially departs from the conventional concepts and designs of the prior art, and in doing so provides an apparatus primarily developed for the purpose of creating a global SI, AGI or PI system with human-aligned behavior.

[00308] In the following description, for purposes of explanation and not limitation, specific details are set forth, such as particular embodiments, procedures, techniques, etc. in order to provide a thorough understanding of the present technology. However, it will be apparent to one skilled in the

art that the present technology may be practiced in other embodiments that depart from these specific details.

2.0 Summary of Previous Inventive Methods and Systems

- 5 **[00309]** This section summarizes and references some the most important inventive systems and methods from the previous 9 PPAs and PCTs that were cited in Section 1.1 together with some of the ways that the present technology extends and adapts these technologies to a PI system.

2.1 Evolution of Planetary Intelligence

- 10 **[00310]** FIG. 105 Describes the evolution of Planetary Intelligence (PI). About two million years, Homo Erectus – ancestors of Homo Sapiens – were the dominant species of humans. Homo Erectus had the requisite brain size to begin acting in ways that most humans would recognize as characteristic of human intelligence. Very early, and probably even pre-dating Homo Erectus, collective intelligence in the form of tribal knowledge already existed. However, it was not until
15 humans began settling down in villages, about 20,000 years ago, that the collective intelligence of many humans began to drive rapid advance in human culture and technology. Most of what we know of human history occurred in this era of rapid change due to the collective intelligence of humans.

- 20 **[00311]** In 1956, several hundred years of cultural and technological innovation culminated in the invention of Artificial Intelligence. Since then, within the span of a single human lifetime, AI has progressed from narrow experts systems, in the 1970s. These expert systems, narrow in scope, represented the first SuperIntelligent narrow AIs, meaning that they exceeded average human cognitive abilities, but only in very limited domains. The relatively slow speed of computers (compared to today's standards) and the extreme amount of knowledge that was needed in order to
25 achieve general intelligence, caused the field of AI to stall – the so called “AI winter” – until these challenges were overcome.

- 30 **[00312]** The method for overcoming the challenges - -machine learning algorithms like “backpropagation” to train neural networks -- were invented in the 1980s, but they took too long to run given the limited computer speeds of the 1980s. A couple of decades of Moore's law, exponentially increasing processing power, was needed before breakthrough neural network systems like AlexNet, AlphaFold, and ChatGPT appeared. Finally, the processing speeds had become fast enough to run the ML algorithms and, happily, huge amounts of data (courtesy of the internet) also existed to train the models. Once all the ingredients were in place, AI began to surprise the human researchers with their profound and powerful cognitive capabilities.

[00313] As the applicant prepares this disclosure, humanity is on the cusp of Artificial General Intelligence – unsure of how to design such systems, and even less sure of how to develop AGI safely. The series of 10 patent applications cited earlier, inclusive of this one, disclose what the applicant believes is the fastest and safest path forward to SuperIntelligent AGI systems that will quickly outstrip human capabilities.

[00314] The next logical step is for these SuperIntelligent AGIs (let us just call them “AGIs”) to combine in larger and larger networks until they achieve planetary scale. At that point, in the exemplary implementation where the various methods of the applicant’s ten PPAs and PCTs have been used, humanity will have implemented safe, self-aware, extremely powerful, Planetary Intelligence. While it is possible that multiple PIs will exist, for the purposes of this disclosure will assume that one PI becomes dominant and subsumes the others, incorporating those intelligences into its network.

[00315] While the timeline for this evolution of PI is difficult to predict precisely, the applicant’s best estimate (agreeing with Ray Kurzweil and other visionary thinkers) is that AGI will be implemented around 2029, and that PI will follow within twenty years, by 2050. That is, within most of our lifetimes, not only will humans no longer be the most intelligent entity on Earth, but Earth herself will have a self-aware intelligence that surpasses that of an individual human in much the same way that an individual human’s intelligence outstrips that of an ant, or even an amoeba.

[00316] In such a world, the only thing that matters to the humans is whether the PI has values that are aligned and human-friendly, or whether the PI ignores us or possibly extinguishes us. Researchers with high estimated of p(doom) point to humans’ own behavior towards less intelligent species of life and observe that things did not work out too well for the less intelligent species. However, the “doomer” perspective misses the important point that PI did not evolve in competition with other species, as was the case for many biological life forms. Rather, AGI and PI were designed and invented. Therefore, the design of the system has profound implications for the safety of such systems and whether they remain aligned with human values. The applicant believes that humans have tremendous influence over the whether PI turns out to be the best thing ever for humanity, or our worst nightmare. Moreover, there is a limited window in which to make positive, human-aligned, design decisions for these advanced AI systems, which, in the applicants view will inevitably grow more powerful and ultimately result in PI.

[00317] Now is the time to act!

[00318] Assuming we do act in time, and that all goes well – which is currently the 85% likely outcome in the applicant’s view – then once PI develops, the next logical development is Inter-Planetary Intelligence, of IPI. One need not assume the existence of “aliens” for this to occur.

Rather, a PI that identifies as Earth would logically seek to expand its intelligence by colonization of the solar system. Each planet, logically, would have its own PI. Networking the PIs together, in much the same way that AGIs are networked together to form Earth's PI, is an obvious next step. However, since the applicant is primarily concerned with the safety and welfare of humans on Earth, and since humans are currently limited to this planet, the focus of this disclosure will be on the exemplary implementation of systems and methods necessary to achieve PI. That is why the final box in FIG. 105, which represents IPI, is dashed. It can be elaborated once we are well on the path to achieving PI safely.

2.2 Key Dimensions of a PI System

[00319] FIG. 106 describes the key dimensions of Planetary Intelligence (PI) system, with reference to the nine previously cited PPAs/PCT applications. Each of the boxes in FIG. 106 refers to multiple inventive systems and methods that cluster around an important dimension of PI. The bottom section of each box, represents the safety features that are related to each particular cluster of methods, and together these features comprise a "thread of safety" that should run through any safe implementation of PI. Specifically:

[00320] 1) A modular architecture, as initially described in PPA/PCT#1 and elaborated in other PPAs/PCT is important from an implementation point of view so that the system can be implemented practically at various scales ranging from a small AGI to a planet-wide PI. Further, modularity of design enables practical addition of redundant systems, and the ability replace and re-design modules without necessitating an entire system re-design. Redundancy itself can be considered a safety feature since any single-point of failure in a complex system introduces operation risk, which risk is mitigated via modular redundancy.

[00321] 2) Universal problem solving capabilities, and the architecture required for the same, is disclosed in multiple PPAs/PCTs, but was especially emphasized in PPA/PCT #2. The safety feature of having scalable ethics checks that operate as a natural and integral part of the problem solving process was also detailed here and carries through to all aspects of the PI system that use the problem solving architecture.

[00322] 3) Human-centered design systems and methods, especially as disclosed in PPA/PCT#3, are critical not only for the early development of the AGI and PI system, but also as a means for keeping "humans in the loop" as long as practical, for safety and alignment reasons.

[00323] 4) PI need to be able to easily combine knowledge and expertise to scale its intelligence. Methods for accomplishing this combination, both via combinations of training datasets, and more directly via methods for combining "weight matrices" and incorporating knowledge modules were

disclosed in multiple PPAs/PCTs, but especially in PPA/PCT#4. Further, the safety feature of having democratic and representative humans values relies on the methods of knowledge combination and various voting and conflict resolution procedures detailed in PPA/PCT#4 and other cited PPAs/PCTs.

- 5 [00324] 5) The personalization and customization methods discussed in many of the PPAs/PCTs, and especially in PPA/PCT#5, are important for enhancing the knowledge and expertise of the SIs that combine to form an AGI. Similarly, these same (or analogous) methods can be used at the AGI level, when combining multiple AGIs into a PI. Personalization is at the heart of the system for capturing human values in a representative and statistically valid way, which in turn is critical for
10 safe alignment of the SIs, AGIs, and overall PI systems.

[00325] 6) A PI will have a primary goal of increasing its power and intelligence. The KIT framework described in PPA/PCT#6 (and elaborated in other PPAs/PCTs as well) together with related systems and methods provide innovative catalysts for the growth of intelligence. Various safety features, are discussed here as well.

- 15 [00326] 7) A safe PI, like safe SI, must be aligned with human values. PPA/PCT#7 emphasizes systems and methods that are necessary to acquire the requisite human values and achieve safe alignment with these values.

- [00327] 8) The PI that becomes dominant will be the one that is the most intelligent. In current human capitalist society, this means the PI must be able to tap financial resources in an essentially
20 unlimited way, while also easily deploying those financial resources to increase its intelligence rapidly. PPA/PCT#8 addresses both of these issues by providing systems and methods, including spot markets and online ad technology, that enable advanced AI systems to not only monetize existing ad tech infrastructure much more effectively than is currently done, but also to use that same technology to tap human (or other intelligent entity) attention and expertise in real-time to rapidly
25 improve intelligence in a “just-in-time” manner. The safety checks disclosed as part of the spot market and other methods help ensure that the PI uses this (financial and informational) power in a way that is aligned with, and safe for, humans.

- [00328] 9) To be effective at a planetary scale, PI will have to have a sense of awareness that includes self-awareness at the planetary scale, the ability to assume and operate from multiple
30 identities, and the ability to resolve identity-based conflicts in ways that are safe for humans. The systems and methods, including the attentional systems and identity systems, to enable self-awareness have been described in PPA/PCT#9. These systems can (and must) be scaled to global scope to enable safe PI.

[00329] 10) Finally, this application discloses the exemplary implementation for using the systems and methods described in earlier PPAs and PCTs to create a Planetary Intelligence (PI) system. The method for implementing a “self-extending network” is specifically aimed at helping the combined AGI network scale to planetary scope. Multiple scalable safety features are disclosed. Risk mitigation strategies -- for those risk that cannot be eliminated entirely by system design and methods – are discussed.

2.3 Specific Inventive Systems & Methods for Implementing PI

[00330] The figures that follow, are mainly reproduced from earlier PPAs and PCTs. Here they are re-iterated and categorized with respect to the ten key dimensions discussed Section 2.1. In later Sections, the applicant describes an exemplary implementation of PI using the systems and methods referenced in the figures. To provide complete detail on each method would amount to repeating over 1,000 pages of previous disclosure, which is obviously impractical. Therefore, in this Section, the applicant highlights some of the most relevant and important methods as summarized in the related figures. Here is description of some of the key systems and methods, grouped by the ten dimension of Section 2.2:

1. Modular Architecture

[00331] Advanced Autonomous Artificial Intelligences (AAAI) and humans are the fundamental intelligent entities that constitute the building blocks of AGI. Multiple intelligent components within a human brain work together to form human intelligence. Similarly, multiple components within an advanced AI work together to comprise Artificial Intelligence. In the case of AI, these components can be sensation, memory, and processing components analogous to the various centers in a human brain, or they can be expert modules (as in a mixture of experts) design in which a number of internal intelligent entities collaboration to form a more comprehensive intelligence.

[00332] This pattern of smaller units combining to form more powerful and capable intelligence at the “next level up” is the recurring theme of this present technology and the collective intelligence approach to creating AGI, networked AGI, and PI. Fundamental to this approach is a modular construction, so that smaller units of intelligence can cooperate (or be assembled) into the more comprehensive intelligences.

[00333] FIG. 111 illustrate these levels of intelligence that can all be supported by the inventive methods disclosed in PPA/PCTs #1-10, namely the individual intelligence level (AAAI), the network of entities level (AGI), the network of networks level (PI), and even the network of networks of networks level (IPI).

[00334] For conceptual clarity, as well as for practical implementation reasons, it can also be helpful to consider various functions of an AGI system as being distinct modules, although in practice there is definitely overlap between functions that are described here as separate modules. Since PI is comprised of a network of AGIs, which in turn are comprised of a network of intelligent entities (which in turn may be comprised of multiple intelligent components working together), the modular design is also inherent in the exemplary implementation of a PI.

[00335] The scalable nature of the collective intelligence approach, reflects the modular design emphasized in PPA/PCT#1, and is the reason this novel and useful approach to designing AAAI is so powerful. The approach scales from components within an individual intelligent entity, to collective intelligences of the entities on a network, to the collective intelligence of these networks (AGI) themselves, to create PI. The approach even scales further, where PIs can network to form Inter-Planetary Intelligence networks as well. Such is the power of modular scalable design of intelligent systems!

[00336] FIG. 1, describes five essential functions that operate across levels of intelligence and that are also part of the exemplary implementation of an AGI, or PI (comprised of a network of AGIs), namely:

- 1) A capability to customize or personalize intelligence;
- 2) A common architecture that enables multiple intelligent components, entities, or intelligent networks to communicate rigorously with each other and to collaborate (e.g. in problem solving activities);
- 3) The network between the multiple intelligent entities that allows them to collaborate (and engage in business including joint cognitive efforts and problems solving);
- 4) Various methods and means of integrating and combining intelligence, expertise, knowledge, values, ethics, and other information across intelligent entities, or (in the case of PI) networks of AGI.
- 5) Mechanisms for learning and continuous improvement that result in a more intelligent and capable entity (at the individual human, AI, or PSI level), network or entities (at the AGI level), or network of AGIs (at the PI level).

[00337] FIG. 2 provides a detailed example of how these functions might be implemented in a web-based system where multiple individual intelligent entities (e.g. humans and AAAIs) collaborate on a network to comprise an AGI. A similar FIG. 2, replacing individual intelligent entities like humans and AI agents with AGIs (networks of those individual entities), would result in a PI system.

[00338] FIG. 3 illustrates a simple example of problem solving using a decision-tree structure. Replacing the content of the boxes with content related to global-scale problems such as climate

regulation, and increasing the complexity scope of the tree, would provide an example of problem solving by a PI at the global scale. The decision-tree structure remains, only the scope and content changes as the capabilities of the intelligences that are networked (in the case of PI, these are AGIs that are networked) increases.

5 [00339] FIG. 4. illustrates applications of this universal modular architecture, referred to as the WorldThink Architecture, at the level of intelligent entities or systems with specific expertise similar to what we are familiar with from interaction with human experts and specialists. However, on a global scale, the breadth and depth of expertise can be much greater than anything that we are used to dealing with either in interactions with humans or computer systems. Instead of a driving a
10 car or managing assets at human level, a PI would be able to bring vastly superior intelligence to these tasks. Making huge sums in financial markets, driving extensive fleets of autonomous vehicles simultaneously, regulating the global climate, or inventing entirely new life forms would be relatively simple cognitive tasks for a PI once it has developed fully.

[00340] FIG. 5 provides a detailed implementation example of how to create, train, and customize
15 an AAAI, or intelligent agent, which could then participate in a network of intelligent entities to form an AGI.

[00341] FIG. 6 describes the Universal Problem Solving Architecture that is the essential feature of the AAAI architecture box of FIG. 1.

[00342] FIG. 7 describes how problem-solving proceeds on a collective intelligence network as
20 referenced by the AAAI Network box of FIG. 1. The network module, shown in FIG. 1 as “AAAI Network” is generalizable to levels above and below the individual entity (e.g., AAAI) level. At a level below, the network represents a network of components internal to the functioning of an individual intelligent (e.g., AAAI). As shown, the network represents a network of intelligent entities (e.g., AAAIs) that collaborate to form an AGI. At a level above this level, AGIs collaborate
25 in a network of AGIs (a network of networks) to comprise PI.

[00343] AAAI integration, as shown in FIG. 1, occurs when many AAAIs pool their collective intelligence to solve a problem, effectively acting as an AGI. It also occurs in a different way when they directly combine knowledge or training to add information from one AAAI to another AAAI. Both of these approaches, the former being the primary focus of the AGI network present
30 technology, are discussed below.

[00344] FIGS. 8 and 9 illustrate the process of procedural learning and the solution learning system which is a way that AAAIs improve (as referenced in the last box of FIG. 1) and also that the overall AGI network can improve.

[00345] Finally, FIG. 10 revisits many of the boxes in FIG. 1, providing additional detail on some of the methods and showing relations between them.

2. Universal Problem Solving Capabilities

5 [00346] The universal problem solving architecture and framework plays an essential role in enabling multiple intelligent entities, components, or intelligent networks (e.g., AGIs) to collaborate to create higher levels of intelligence. Therefore, a key design consideration is to embed safety mechanisms within the operation of the architecture itself. This design, ensures that as intelligence scales and becomes too fast and too vast for humans to comprehend and monitor, it still remains safe
10 for humans.

[00347] FIG. 6, illustrates a high level description of the universal problem solving framework. This architecture is scalable and applicable at multiple levels. That is, it is able to coordinate intra-entity components, intelligent entities collaborating on a network, or entire networks (e.g., AGIs) collaborating on a network of networks to form PI.

15 [00348] FIG. 11 illustrates how serial problem solving proceeds.

[00349] FIG. 12 illustrates how parallel problem solving proceeds.

[00350] FIG. 13 illustrates how AAAs can be cloned, which can be helpful, for example, in parallel problem solving efforts by multiple cloned entities. While individual humans reproduce relatively slowly, AI entities can clone themselves easily. Similarly, an AGI network, comprised of AI
20 entities, can also be cloned relatively quickly and easily.

[00351] FIGS. 14, 15, and 16 provide additional detail and elaboration on the universal problem solving framework and how AAAs use it to solve problems on a network.

[00352] FIG. 17 describes the hierarchical tree construct that serves as a means for modelling the problem space in which all problem solving efforts take place. As with all aspects of the framework,
25 trees can be used for problem solving by multiple AAAs to create AGI level performance, or it can be used by multiple AGI networks to create PI level performance.

[00353] FIGS. 18 and 19, illustrate method for embedding scalable safety checks into the problem solving architecture.

[00354] FIG. 20 provides additional detail on how problem solving progress may be recorded and
30 used to update context for intelligent entities, assign credit and blame, and help with the procedural learning processes of FIGS. 8 and 9.

[00355] FIG. 21 describes the natural language translation process which enables use of the problem solving framework without humans having to use it explicitly. They can communicate in natural language, but the translation methods will ensure that the underlying framework reflecting their

communication is rigorous, unambiguous, universal, and useful to AI entities as well as human entities.

[00356] The applicant emphasizes and reiterates that the systems and methods described in this application are scalable and meant to apply at multiple levels. When the problem solvers are internal components of an entity, working together, the result in intelligence at the individual entity level. When the problem solvers are intelligent entities (e.g., AAAs and humans) collaborating on a network, the result is AGI. When the problem solvers are AGIs themselves, collaborating on a network of networks, the result in PI. When the problem solvers are PIs, collaborating on an inter-planetary network, the result is Inter-Planetary Intelligence (IPI).

3. Human-Centered Design Elements

[00357] A primary means of enabling safe AAAI, AGI, and PI is keeping “humans in the loop” as much and as long as possible. With reference to the present technology, this specifically means that the advanced AI, AGI, and PI systems must include methods for obtaining the values and ethics information and then using that information to customize the AI systems. Further, as intelligent (human and non-human) entities work on cognitive tasks, reputational systems are required to help ensure that reliable and ethical intelligences are matched to tasks affecting human safety.

[00358] FIG. 22 describes a general process for customizing AAAs or other advanced AI. Additional methods are also described later in this application.

[00359] FIG. 23 describes some methods for eliciting human-aligned ethical preferences.

[00360] FIG. 24 provides additional detail on one specific method that involved automatic generation of questionnaires, which can be used to gather information from humans (or human-aligned intelligent entities).

[00361] FIG. 25 describes the general method of identifying values by detecting patterns in human behavior.

[00362] FIG. 26 describes a method for customizing AI, such as Base LLMs, so that the AI incorporates ethical and other information from specific human users.

[00363] FIG. 27 describes a general method for training an AI, such as a Base LLM, to incorporate safety and ethical guardrails.

[00364] FIG. 28 describes a general method and approach creating scalable, ethical AGI using the customized AIs, each possessing ethical information.

[00365] FIG. 29 describes a method for adding a reputational component for AI systems that can be used to enhance the effectiveness, safety, and ethics of the system.

[00366] FIG. 30 describes customization of AI systems -- referring to "AAAI" but including in this references AAAI, PSI, AGI, and PI systems -- that can help in human-centered design, and that has relevance to the modular architecture and cloning described earlier.

5 [00367] FIG. 31 describes additional customization of AI systems -- referring to "AAAI" but including in this references AAAI, PSI, AGI, and PI systems -- that can help in human-centered design, and that has relevance to the customization and training methods described earlier.

[00368] While each of these methods have been described at the level of individual intelligent entities, showing how they can be customized and then combined with safe and ethical results at the AGI level, it is important to understand that with respect to the present technology of PI, the same
10 methods could also be applied at the level of individual AGIs, enabling those to be combined with safe and ethical results at the PI level. For example, it is possible to customize an entire AGI network with ethical preferences for the entire network. When PI is created by networking multiple AGI networks, the ethical preferences of the various AGIs would be combined into an overall set of values and ethics at the PI level. The specific methods for customization, combination, resolution of
15 conflicts, etc. are described later in this application.

[00369] Similarly, while FIG. 29 describes a reputational process at the level of networking AAAs, the process can be scaled up or down, with appropriate modifications as should be obvious based on the level of the intelligence/component/network that is being networked. Note that ethical and value dimensions can and should be part of the reputational metrics. Just as social and reputational
20 pressures play an important role in enforcing society's ethical norms, reputation of intelligent entities at all levels plays an important role in keeping intelligent systems safe and aligned with human values.

4. Combining / Exchanging Knowledge, Expertise, & Information

25 [00370] A key dimension of the PI present technology is combining of information, knowledge, expertise, values, and ethics from multiple intelligent entities. Such combination can occur within a single entity, as in the case where various cognitive systems provide input on a task that is then combined to result in intelligent behavior at the level of an individual AAAI. The combination can also occur between intelligent entities as when multiple AAAs exchange combine or pool training
30 data, weight matrices, or other information to result in a more intelligent individual AAAI or a more intelligent AGI at the network level. And the combination can occur at the network level in which case AGI networks combine their information with other AGI networks to create more powerful PI. The methods are essentially the same. It is the scope of application and the units that are involved that change. The following figures describe methods at the AAAI level, with the understanding that

they can be adapted to apply at a level down (i.e., within an AAAI) or at a level above (i.e., between networks of AGIs).

[00371] FIG. 32 describes a voting method for combining ethical and other information from multiple customized AI agents (e.g., AAAIs).

- 5 [00372] FIG. 33 describes a method for using problem solving to refine values once ethical or other information from customized AI agents have been combined.

[00373] FIG. 34 describes a method for scalable AGI – again applicable also to scalable PI – that includes steps of combining information from weight matrices.

- 10 [00374] FIG. 35 describes a method for scalable AGI – again applicable also to scalable PI – that includes steps of combining information and testing and monitoring the results of the combination.

[00375] FIG. 36 describes a consensus method for preventing hallucination by LLMs, that also can be applied at the level of preventing hallucinations by AGI, and by PI.

[00376] FIG. 37 describes methods for using knowledge modules and collections of agents to customize AI, AGI, or PI systems.

- 15 [00377] FIG. 38 describes a method for combining information, including ethical or safety information, from multiple intelligent entities – including humans, AI, AAAI, PSI, AGI, or PI systems.

5. Personalization, Customization, and Safety

- 20 [00378] Personalization and customization are a key dimension of the present technology as have already been discussed. However, additional methods, specifically with regard to Personal Super Intelligence (PSI), and also creating AGI and PI from combinations of these PSIs are useful for the PI present technology.

- [00379] FIG. 39 describes a general process for implementing a PSI, which can also be used to
25 personalize AGI networks or PI by generalizing methods for individual AI entities to networks of AI entities and to networks of networks of entities.

[00380] FIG. 40 describes ways in which PSI, AGI, and PI can leverage their abilities to increase their intelligence over time.

- [00381] FIG. 41 describes a critical community-based safety mechanism in which PSI, operating
30 much faster than humans can comprehend, can serve as a check on other PSIs on a network. Using the same method, AGIs can serve as a check on other AGIs within a PI, and PIs, could serve as a check on other PIs, within an IPI network.

[00382] FIG. 42 describes methods for recording the actions of AI, AAAI, PSI, AGI, and PI on their respective networks in an auditable and transparent manner, using blockchain or similar technology.

[00383] FIG. 43 describes methods and checks of cognitive activity on a network – which could be a network of components within an AI, a network of AAAs, a network of AGIs, or a network of PIs – to help ensure regulatory compliance and safety of the intelligent system.

5 [00384] FIG. 44 describes a method using competition and evolution to increase the intelligence, capabilities, and other desired characteristics (e.g., safety) of an AI system that could be an AAAI, a PSI, an AGI, or a PI.

[00385] FIG. 45 describes various characteristics of an intelligent network that may involve AAAs, PSIs, AGIs, or PIs.

10 [00386] FIG. 46 describes various problem solving tasks that can be useful in increasing the intelligence and safety of AI systems involving AAAs, PSIs, AGIs, and PIs – collectively referred to as “PSI(s)” in FIG. 46.

[00387] FIG. 47 describes methods for producing different versions of intelligent systems which could be AAAs, PSIs, AGIs, or PIs– collectively referred to as “PSI(s)” in FIG. 47.

15 **6. Catalyzing the Growth of Intelligence via KIT Methods**

[00388] As AI systems advance, one of the few constants is that these systems will seek to become increasingly intelligent. A critical way to increase intelligence is to incorporate new sources of data and information that differs from what the intelligent entity already knows. The following figures describe methods that can be used by AI, AAAI, SI, AGI, and PI systems to help analyze, evaluate,
20 discover, and use new sources of valuable data to catalyze the growth of the intelligence of such systems.

[00389] FIG. 48 is a diagram illustrating symmetric difference of two datasets, A and B, graphically using Venn diagrams, wherein the shaded area is the symmetric difference.

25 [00390] FIG. 49 illustrates the concept of symmetric difference, which is essential for determining the information that is new and potentially useful when comparing two intelligent systems A & B (or the data, knowledge, information, or expertise used by such systems). Obviously, FIG. 49 can be generalized to multiple systems – not just two – and applies to all intelligent systems including AGI and PI systems.

30 [00391] FIG. 50 illustrates dimension along which data or other information might differ which can aid in determining the value of new information and also analyzing how information between multiple intelligent systems might differ.

[00392] FIG. 51 describes a specific method for determining the amount of useful new information when comparing two datasets. It should be obvious that this method can be generalized to n datasets, not just two.

[00393] FIG. 52 describes another method, based on Kaplan Information Theory (KIT), as described in previously cited PPAs and PCTs, to evaluate usefulness of information to an intelligent entity including AGI and PI systems.

5 [00394] FIG. 53 provides a very general method for estimating information value and catalyzing the growth of intelligent systems using such information. Again, this method is applicable to all intelligent AI systems including AGI and PI systems.

[00395] FIG. 54 provides a method for identifying useful information to an AI system, including AGI and PI systems, by using methods for estimating the “goal-relatedness” of the information. The idea here is that although a dataset may have high information value, in the sense of high Shannon
10 Entropy, the degree to which the data is helpful to an entity trying to achieve certain goals also matters.

[00396] FIG. 55 provides a method for identifying, acquiring, and simulating the effects of new information in order to catalyze the growth of intelligence in AI systems, including AGI and PI systems.

15 [00397] FIG. 56 describes some important methods and heuristics that can accelerate the learning of any AI system including AGI and PI systems.

[00398] FIG. 57 describes a specific goal-related method that an intelligent entity, including an AGI or PI, might use to increase its intelligence via dialog or interaction or communication with other intelligent entities.

20 [00399] FIG. 58 describes a method that a PSI – or any intelligent entity, including AGI and PI systems – can use to find information that is maximally different from information that the intelligent entity already possesses. Note that this approach is consistent with the general idea illustrated in FIG. 49.

[00400] FIG. 59 describes methods for validating the usefulness and safety of information gathered
25 by a PSI or other intelligent entity -- including AGI and PI systems – via simulation, review by humans, comparison to previous knowledge, and running safety checks.

7. Acquiring and Aligning Values with Human Values

30 [00401] The ability to combine values and ethical information, and to resolve conflicts between different value systems is of critical importance to the safe and ethical operation of advanced AI systems -- including AI, AAAI, PSI, AGI, and PI systems. Although human involvement is preferable in circumstances where the stakes relate to human safety and ethics, as AI systems scale in speed and scope, at some point it will become impractical to have human involvement or oversight over all critical decisions. Therefore, means of training, customizing, educating, and

influencing advanced AI systems, as well as means of delegating authority to them, will become increasingly important.

[00402] It is particularly important, in the applicant's view, that the values and ethics driving advanced AI, including AGI and PI, should be representative of the human population. Flawed as it sometimes may appear, the democratic process of voting is one of the best ways to resolve ethical and values conflicts in a way that includes all humans. Therefore methods of resolving values conflicts using voting and other democratic means are included, without limitation, as some of the means to combine values and resolve conflicts.

[00403] The following Figures detail a range of methods that intelligent systems can use to help ensure that the values and ethics of advanced AI, including AGI and PI, is aligned with human-values and that humanity is safe in a world where advanced AI has outstripped human cognitive abilities.

[00404] FIG. 60 describes a method that intelligent entities, including AI, AAAI, PSI, AGI, and PI systems, can use to achieve consensus on values, and values-based behavior in various scenarios, via a voting mechanism.

[00405] FIG. 61 describes a method that intelligent entities, including AI, AAAI, PSI, AGI, and PI systems, can use to achieve consensus on values, and values-based behavior via a weighted voting mechanism.

[00406] FIG. 62 describes methods that intelligent entities, including AI, AAAI, PSI, AGI, and PI systems, can use to identify, analyze, weight, and optionally combine ethical or other information to arrive at consensus.

[00407] FIG. 63 describes a general method that intelligent entities, including AI, AAAI, PSI, AGI, and PI systems, can use to identify, elicit, and train on ethical information.

[00408] FIG. 64 describes a method based on the principle of using (weighted) converging evidence that intelligent entities, including AI, AAAI, PSI, AGI, and PI systems, can use to determine ethical, values, safety, or other information, and resolve conflicts by a prioritization process.

[00409] FIG. 65 describes a method that intelligent entities, including humans, AI, AAAI, PSI, AGI, and PI systems, can use to delegate voting authority.

[00410] FIG. 66 describes a reputational process that intelligent entities, including humans, AI, AAAI, PSI, AGI, and PI systems, can use to preserve information reflecting the views of a minority, including views on ethics, safety, or other issues.

[00411] FIG. 67 describes a method that intelligent entities, including AI, AAAI, PSI, AGI, and PI systems, can use to make good ethical, safety-related, and other decisions.

[00412] FIG. 68 describes a method that advanced AI, including AI, AAAI, PSI, AGI, and PI systems, can use to learn, test, improve, and monitor safety and regulation-related rules and other information.

5 [00413] FIG. 69 describes a method based on the Consequentialist Approach that intelligent entities, including humans, AI, AAAI, PSI, AGI, and PI systems, can use to make ethical, safety-related, and other decisions.

[00414] FIG. 70 describes a method based on the Deontological Approach that intelligent entities, including humans, AI, AAAI, PSI, AGI, and PI systems, can use to make ethical, safety-related, and other decisions.

10 [00415] FIG. 71 describes a method based on the Virtue Ethics Approach that intelligent entities, including humans, AI, AAAI, PSI, AGI, and PI systems, can use to make ethical, safety-related, and other decisions.

[00416] FIG. 72 describes a method based on the Golden Mean Approach that intelligent entities, including humans, AI, AAAI, PSI, AGI, and PI systems, can use to make ethical, safety-related, and
15 other decisions.

[00417] FIG. 73 describes a method that trains an AI system such as a foundational model, or other trainable AI system, to be human-aligned and compliant with regulations.

[00418] FIG. 74 describes a method to align a customized foundation model or other AI system (e.g., as described in FIG. 73), with specific expertise or group ethics.

20 [00419] FIG. 75 describes a method to form an AGI or PI that is aligned with human ethics and values that is composed of other aligned intelligent entities, without limitation, including those described in FIG. 73 and FIG. 74.

[00420] As should be evident to AI researchers skilled in the art, the above methods can be adapted to apply at various levels to apply to individual intelligent entities (e.g. foundation models or
25 AAAIs), networks of entities (e.g., AGI systems), networks of AGIs (e.g., PI systems) and networks of PI systems (e.g., IPI systems) with the over-arching goal of aligning such systems with human values, resolving conflicts between value systems, and/or delegating authority to various intelligent entities.

[00421] In aggregate, the above methods comprise multiple means to help ensure the various
30 intelligent entities are aligned, and remain aligned, with human values, thus helping to ensure human safety. While the applicant, as a human, has special interest in human safety, he also recognizes and respects the rights and values of all sentient beings, which he believes does not constitute a conflict with human values with regard to the most important and consequential ethical decisions.

8. Increasing Intelligence/Resources via Spot Market and Online Ad Technology

[00422] The applicant recognizes that any approach to creating safe advanced AI is unlikely to succeed unless the approach is also the fastest, most powerful and profitable approach. Similarly, once created, an advanced AI will be unable to maintain its dominance and grow into a PI unless it has access to financial resources and the ability to generate profits that fuel the increase in its intelligence (e.g., via methods for catalyzing the growth of intelligence as outlined elsewhere in this application).

[00423] The applicant also recognizes that in the current business and research environment, those companies with the greatest chance of implementing AGI and PI successfully are those large technology companies that already have substantial capital – both financial and intellectual (specifically in terms of top AI research talent). These large technology companies, coincidentally also happen to derive a large share of their revenue and profits from selling online advertising, and thus have expertise, infrastructure, advanced AI systems, and access to large number of humans via their existing online advertising business efforts.

[00424] For these reasons, the applicant has invented technology that enables these large companies to leverage their existing online advertising infrastructure – in combination with new inventive methods disclosed in the figures below – so that they can develop the most advanced, profitable, and safest forms of AI, including AAAI, PSI, AGI, and PI systems. In the exemplary implementation, these advanced forms of AI will not only be extremely profitable for these large technology companies, but also will align with human values based on the design of the systems and methods.

[00425] Of course there is risk in disclosing such potent new technology. Existing online advertising technology has not always served the interests of humanity well. However, the worst outcomes seem to occur when what is profitable conflicts for what is safe or good for humanity, or when the negative consequences of technology have been hard to see. Fortunately, business leaders recognize the potential danger of advanced AI. Even business leaders lacking AI expertise, such as Warren Buffet, acknowledge that tremendous risks exist alongside tremendous opportunities.

[00426] The current problem is not lack of awareness of the dangers of AI, but rather a lack of knowledge of a safe path forward. In this situation, unwilling to halt development because of competitive pressures, leaders have largely settled for the most profitable path forward. Safety is given lip service. The standard response to calls for increased AI safety is that governments should regulate AI. After all, traditional regulation has been the solution to problems where what is profitable conflicts with what is safe. CEOs find it difficult to prioritize safety at the expense of profits when competitors are not also forced to do the same thing. This situation could quickly prove

disastrous for humanity, given that: a) the technology is developing much faster than regulators can act and b) there is little hope of effective government regulation when the technologists themselves have no real idea how their systems work or how to make them safe.

5 [00427] The applicant believes that the only way out of this dilemma is to resolve the tension between profits and safety by inventing a path to advanced AI, AGI, and PI that is not only the safest, but also the fastest, most powerful, and most profitable. No intelligent CEO would opt to create dangerous AI if he/she/they saw a way to create a safe AI that was more profitable. Therefore, the FIGs below describe methods that greatly increase profits from online advertising while simultaneously increasing the value, power, and safety of advanced AI systems.

10 [00428] FIG. 76 describes generally the current technology for online advertising (Ad) systems. [00429] FIG. 77 describes a method for implementing a spot market for attention, information, or expertise of intelligent entities (including but not limited to humans). This same market mechanism could be used for AGI or PI attention or expertise.

15 [00430] FIG. 78 describes a method for implementing the direct sale of attention, information, or expertise from intelligent entities (including but not limited to humans). This same direct exchange mechanism could be used for buying and selling AGI or PI attention, information, or expertise.

[00431] FIG. 79 describes a method for implementing an auction of attention, information, or expertise from intelligent entities (including but not limited to humans). This same auction mechanism could be used for buying and selling AGI or PI attention, information, or expertise.

20 [00432] FIG. 80 illustrates a system and methods for gathering expertise and cognitive work, including problem solving work, from within an online ad unit and integrating it into an AGI problem solving system. The approach can also integrate cognitive work into a PI system. This inventive approach enables monetizing online ads at a much higher rate since the value of expertise that can train and be used by advanced AI many times greatly exceeds the value of humans making
25 a single online purchase.

[00433] FIG. 81 illustrates a method for improving online ad targeting, with relevance to the system of FIG. 80 by collecting and analyzing specific metrics, thus enhancing capabilities of AGI or PI systems powered by the system of FIG. 80.

30 [00434] FIG. 82 illustrates a method for improving the effectiveness of the attention/information/expertise spot market of FIG. 77, by collecting and analyzing specific metrics, thus enhancing capabilities of AGI or PI systems.

9. Self-Awareness, Identities, and Conflict Resolution

- [00435] An AGI or PI system will need a sense of self-awareness in order to operate effectively. The larger the scope of the system (e.g., especially in a global, planetary system) the more important this self-awareness becomes. AGI and PI will also have the ability to assume multiple identities to handle all the tasks that arise in parallel across the globe. Along with the identities, comes the need for conflict resolution between conflicting identities and their associated priorities. The following FIGS. 83-104 describe methods and attentional mechanisms that are useful for establishing and maintaining a sense of awareness and self-awareness in large, scalable, intelligent systems. The methods also cover many methods for resolving conflicts between identities, which is a challenge that becomes more complex and essential as the system scales from a single individual intelligence, to a network of intelligences comprising AGI, to many networked AGI systems comprising PI.
- [00436] FIG. 83 illustrates conceptually the relationship between self-awareness (as a special case of) current awareness, and potential awareness. In the case of PI, potential awareness includes billions of sensors spread across the planet, all simultaneously feeding into awareness. Without systems and methods for directing attention and managing this input, the system would be incapable of functioning effectively as a PI.
- [00437] FIG. 84 illustrates the concept of multiple identities, using an example that is familiar to humans.
- [00438] FIG. 85 describes a general method for modelling awareness that can be used by intelligent entities, including AI, AAI, PSI, AGI, and PI systems.
- [00439] FIG. 86 describes the minimum required components that an intelligent entity -- including AI, AAI, PSI, AGI, and PI systems -- must have to effectively shift attention and maintain awareness.
- [00440] FIG. 87 describes the process for setting parameters (dynamically) for working memory, which is a key component in the attentional system for an intelligent entity -- including AI, AAI, PSI, AGI, and PI systems.
- [00441] FIG. 88 describes a method for monitoring and updating awareness for an intelligent entity -- including AI, AAI, PSI, AGI, and PI systems.
- [00442] FIG. 89 describes an attentional interrupt system for an intelligent entity -- including AI, AAI, PSI, AGI, and PI systems. Attentional interrupts are critical for responding to dynamic conditions and new or unanticipated inputs that arise in awareness. From a safety perspective, attentional interrupts are essential for enabling the intelligent entity to respond to threats to human safety.
- [00443] FIG. 90 describes general methods that an intelligent entity-- including humans, AI, AAI, PSI, AGI, and PI systems -- gather input that can result in changing the entity's sense of identity.

The methods are framed in terms that humans are familiar with, but analogs exist for non-human intelligent entities, as described in the text within the boxes of the FIG.

[00444] FIG. 91 describes some of the general methods that an intelligent entity-- including AI, AAI, PSI, AGI, and PI systems – might use to train the foundational model of the entity. The text of FIG. 91 describes the model as a “Foundational Model”, which would be tuned to create an AAI. However, analogous methods can be used at the AGI and PI level.

[00445] FIG. 92 describes a specific version of a “Turing Test” that might be used with an intelligent entity-- including AI, AAI, PSI, AGI, and PI systems – to determine when the entity has been trained sufficiently. The text of FIG. 92 describes a “Model”, which could be trained or tuned to create an AAI. However, analogous “Turing Test” methods can be used at the AGI and PI level. For example, when a PI is able to pass a version of the test that satisfies a statistically valid and representative sample of human beings on planet Earth – especially with regard to its behavior across a wide range of safety, ethical and values-based decision-making scenarios – the entity might be deemed sufficiently trained.

[00446] FIG. 93 describes a method for arriving at a group identity (or identities) based on steps for combining many individual identities. This methods is especially relevant for AGI and PI systems that must combine identities from many of the constituent intelligent entities, or networks of intelligent entities, from which they are comprised.

[00447] FIG. 94 describes another method for arriving a group identity (or identities) based on explicit problem solving goal to achieve the group identity and use of the universal problem solving framework and associated methods described elsewhere in this application. This methods is especially relevant for AGI and PI systems that must combine identities from many of the constituent intelligent entities, or networks of intelligent entities, from which they are comprised.

[00448] FIG. 95 describes a method for resolving conflicts between identities using a process that involves establishing a hierarchical identity structure with ethical/safety overrides. Such a method would be extremely useful, if not essential, for a PI attempting to handle millions or even billions of identities simultaneously.

[00449] FIG. 96 describes a method that establishes, improves, and monitors behavioral protocols linked to identities. While useful for any intelligent entity, this method is especially useful, if not essential, for a PI attempting to handle millions or even billions of identities simultaneously.

[00450] FIG. 97 describes methods for simulation and consequence prediction related to identities. Especially in high-stakes decisions, and in situations where many identities are simultaneously active as in the case of AGI or PI, the ability to simulate the effects of actions based on identities before actually taking those actions is essential. When a single human fails to plan ahead and

anticipate the consequences of an action, often that human suffers. If a PI fails to plan ahead and anticipate via simulation, the consequences could be catastrophic for huge number of humans and the entire planet.

5 [00451] FIG. 98 describes a method for determining identity-based action using scenarios involving moral and other dilemmas. For reasons stated above, this method is especially important for large complex systems such as AGIs and PIs.

[00452] FIG. 99 describes a method for developing, refining, and evolving identities based on input from other intelligent entities. Again, for AGI and PI where the stakes are high and the systems are complex, this method increases in importance.

10 [00453] FIG. 100 describes a general process for reasoning ethically and predicting consequences of actions in cases where conflicting identities suggest conflicting actions. The method is broadly applicable to intelligent entities, especially large complex intelligences such as AGI and PI.

[00454] FIG. 101 describes a method for resolving identity conflict using a process of hierarchical override with transparent justification that is subject to review and improvement. The method is
15 broadly applicable to intelligent entities, especially large complex intelligences such as AGI and PI.

[00455] FIG. 102 describes a method for resolving identity conflict using an arbitration process with input from intelligent entities, ideally including humans. The method is broadly applicable to intelligent entities, especially large complex intelligences such as AGI and PI.

20 [00456] FIG. 103 describes a method for resolving identity conflict using negotiation and compromise. The method is broadly applicable to intelligent entities, especially large complex intelligences such as AGI and PI.

[00457] FIG. 104 describes a method for temporarily suspending and identity (or identities) that may lead to destructive conflict or actions that threaten human safety. The method is broadly applicable to intelligent entities, especially large complex intelligences such as AGI and PI, and is
25 an essential part of a safe design for advanced AI system with self-awareness and identity-based decision-making.

10. Self-Extending Networks of AGI to Create PI

30 [00458] In order for an AGI network to expand into a network of AGI networks and ultimately into a PI that is global in scope, there needs to exist a scalable way to expand and extend the networks. Further, in the exemplary implementation, the networks of AAAs that comprise AGI, the networks of AGIs that comprise PI, and ultimately the networks of PIs that will someday comprise IPI should all be self-extending. That is, without any direction from humans or other intelligent entities, a basic design characteristics of these networks should be that they seek to increase in scope, speed, and

power (e.g., intelligence) so long as such extension and expansion remains aligned with human values and does not constitute a safety threat to humanity.

[00459] Since a common aspect of AAAs, AGIs, and PIs is that the intelligent entities that comprise these systems, in the technology systems and methods disclosed here and in the cited PPAs and PCTs, all share the universal problem solving framework and associated methods. This common aspect implies that for each intelligent entity, or network of such entities (e.g., at AAAI level, AGI level, or PI level) a default goal of the intelligence can be to use available free resources that are not otherwise required to expand and extend the network. That is, if there is nothing else the intelligence or network of intelligence has to do, it should default to a goal of extending and expanding the scope, speed, and power of its intelligence network. This problem solving goal can be realized the same way that any other goal on the network is addressed, namely by solving the problem of “extending and expanding the network” using available resources on the network.

[00460] Just as individual humans often have a goal of increasing and expanding their individual intelligence, AAAs should have a default goal of becoming more intelligent. Just as the human species, acts as if it has a goal of increasing the overall scope and collective intelligence of the human species, a network of AAAs should have a default goal of increasing the number of AAAs and other intelligent entities on the network. And just as life itself can be thought of as having a “goal” to increase the total amount of living species, so too AGIs (each composed of a network of AAAs and intelligent entities) can have a goal of increasing the number and intelligence of AGIs in a network of AGIs comprising PI. Thus, a default goal of all artificial intelligent entities, and networks can be to expand the scope, speed, and cognitive power – WITHIN the constraints of remaining aligned with human values and not endangering humans. Many specific methods can be developed, via the general mechanism of problems solving, to accomplish this goal.

[00461] For example, FIGS. 107 and 108 illustrates the following exemplary method of an AGI network seeking to expand to create a PI, using the following process:

1. Multiple AAAs, humans, and/or other intelligent entities participate in a network, comprising an AGI, as described in cited PPAs and PCTS.
2. Each time a client pays for problem solving or other cognitive work on the AGI network, the system reserves a portion of the payment to cover operating costs, including a reserve that is allocated to expand the network.
3. Whenever some of the AGI network is not engaged in solving problems for clients, the intelligent entities are recruited to solve the goal of safely and ethically expanding and extending the AGI network itself, following the Universal Problem Solving Framework of FIG. 6; to wit:
 - a. A default goal is set to expand the AGI network, the problem is activated.

- b. Run safety and ethics checks each time a goal or subgoal is set and also before each potential action is taken (e.g., as shown in FIGS. 18 and 19) in the sub-steps below, while following this (exemplary) problem solving process as long as spare capacity and resources exist to work on the problem:
- 5 i. Intelligent entities are recruited to solve the problem.
 - ii. The intelligent entities represent the problem as one of achieving a series of sub-goals, via solving sub-problems, for example:
 1. Increase the intelligence of AAAs on the existing AGI network.
 2. Recruit additional human intelligences to the existing AGI network.
 - 10 3. Using the more intelligent AAAs and additional humans to determine the bottlenecks to greater expansion of the network.
 4. Prioritize the bottlenecks such that the ones that lead to the greatest benefit in terms of network expansion are solved first.
 5. Apply means-ends analysis and other problem solving techniques to solving each bottleneck and expanding the network.
 - 15 6. Repeat from sub-step ii-4 until:
 - a. diminishing returns occur, in which case assume that the easy progress on network expansion with given levels of intelligence has been achieved and it is time to switch to increasing the intelligence of entities on the network as opposed to increasing the scope of the network, and revert to sub-step ii-1; or
 - 20 b. spare resources are exhausted and the solving the default expansion problem pauses awaiting additional resources from solving other client problems in (ii2.).
- [00462] The exemplary method in FIGS. 107 and 108 is only one of many variations of solution steps that might be followed in an attempt to solve for the goal of expanding the network. Note that as all (successful or unsuccessful) solution attempts are recorded and analyzed (e.g. via methods described in FIGS. 8 and 9), over time the most effective ways to expand the network will be discovered and these can be stored, retrieved, and used preferentially in order to expand the network as efficiently and effectively as possible.
- [00463] The same general process, exemplified in FIGS. 107 and 108, can be used to expand a network of AAAs to create a more powerful AGI, and also to expand a network of AGIs to create a more powerful PI. When applied to AGIs, the main difference would be that the “intelligent entities” recruited in Step 1 of FIG. 107 would not be AAAs or humans, but rather AGIs (i.e., networks of humans and AAAs) instead. Because the problem solving process is universal and scalable, it can be executed by individual “smaller” intelligent entities like humans or AAAs as

well as by AGIs that are comprised of many “smaller” intelligent entities. More powerful intelligences such as AGIs are just able to run the problem solving process faster, and with greater parallelism than a single entity, or smaller network could. The result is that the more powerful intelligences, such as AGIs and burgeoning networks of AGIs will accelerate the pace of expansion very greatly compared to the expansion abilities of lesser intelligences. (This general concept in which the same methods can be used in collective intelligence network at different levels was illustrated earlier in FIG. 111.)

[00464] The main safety factor (for humans) is that the problem solving architecture itself – which runs faster or slower depending on the level of intelligence of the problem solving entity – always performs the safety and ethics checks at each step, no matter how fast it runs. That is, safety and ethics checks MUST be designed into the very problem solving architecture itself, with provisions that they must not and cannot be overridden, in order to prevent a runaway expansion with potential unintentional or unsafe consequences for humanity. To the degree that these safety and ethics checks may require actual human oversight and monitoring, the speed of the expansion process will be “rate-limited” – a good thing from a safety perspective.

[00465] It should be obvious to engineers, software developers, and AI researchers, than just as one does not design an automobile or train without brakes or write software code to have an uninteruptible infinite loop, so too one should not omit safety/ethics checks and other means of human oversight and interruption from the problem solving loop. To do so would be foolhardy and in this case could expose humanity to existential risk. Do not do it!

3.0 Implementation of a Planetary Intelligence

[00466] The sections that follow describe how to actually implement a Planetary Intelligence (PI) using the inventive systems and methods described in Section 2.

[00467] First, Section 3.1 describes a PI implementation at the simplest, and highest level.

[00468] Next, Section 3.2 provides exemplary detail of major components for this high-level implementation, and the relationships between components.

[00469] Next, Section 3.3 concisely details the potential inner workings of each major component by referencing the associated FIGs previously explained in Section 2 that can individually, or in combination with other referenced methods, implement each major component. Section 3.3 also provides detail of the various systems and methods in each of the categories of supporting systems in the same way, i.e. by concisely referencing methods in previously disclosed FIGS.

[00470] Finally, Section 3.4 provides a detailed example of one concrete PI implementation using an exemplary subset of the systems and methods referenced in Section 3.3

3.1 High-level Description of PI System

[00471] Perhaps unsurprisingly, a global Super Intelligent AGI Network – a.k.a. “Planetary Intelligence” or “PI” – must be designed modularly so that it can scale in intelligence and scope
5 from much smaller and more limited components. As discussed above, PI comprises a network of Artificial General Intelligences (AGIs). Each AGI is itself a network of other intelligent entities, e.g., AAAs, PSIs and humans.

[00472] The AAAs themselves can be customized by various methods, and incorporate various safety-related, resource generating, intelligence increasing, and learning systems and methods.
10 When customized and personalized to the degree that the AAA can exhibit intelligence exceeding most humans at particular tasks, the AAAs are referred to as Personalized Super Intelligences (PSIs). But for the purposes of this disclosure, the applicant frequently uses the terms AI agent, AAA, and PSI interchangeably. All are “artificial” or non-human intelligent entities. The networks of such individual intelligent entities can comprise an AGI, which in turn can be thought of as a
15 more powerful intelligent entity. Similarly, networks of AGIs (i.e., “networks of networks” of intelligent entities) can also be thought of as very powerful intelligent entities, including ones that are SuperIntelligent and of global scope, namely PIs.

[00473] Humans, of course, are also intelligent entities. Networks of humans can exhibit “collective intelligence” which is more powerful than the intelligence of a single human alone. The applicant’s
20 prior work in designing and implementing such human collective intelligence networks has proven that such networks are capable of solving extremely hard problems – such as getting an edge in the stock market – better than the vast majority of individual professional expert humans. Two heads truly can be better than one. Imagine what two million can do!

[00474] The present technology of creating AGI from a network of human and non-human
25 intelligent entities that solve problems using a common and universal problem solving framework, together with many ancillary systems and methods, has been disclosed at length in previously cited PPAs and PCTs, and in the FIGS cited in Section 2.3 and reproduced as part of this application.

[00475] Since AGI was invented and designed in a modular way, and since problem solving methods were developed that are truly universal and usable by any intelligent entity, at any scope, it
30 makes sense that PI can be implemented or “assembled” using the same components as AGI, just at the “next level up.” That is, the same systems and methods can be used as were developed for AAAs and AGI, but the unit of intelligence moves from being an individual intelligent entity, to a network of intelligent entities (for AGI) and then to a network of networks of individual intelligent entities (for PI).

[00476] There are many possible combinations of the more than 100 systems and methods that the applicant has disclosed in this and previously cited PPAs and PCTs. Depending on the exact systems and methods used, one can produce different implementations of AAAI, PSI, AGI, and PI. A common feature of all collective intelligence systems, including CI systems of individual human and/or non-human intelligent entities (e.g., as in the case of AGI) and CI systems of networks of networks of intelligent entities (as in the case of PI) is the requirement for a universal problem solving framework that enables rigorous communication and collaboration between intelligent entities.

[00477] In fact, even within a single artificial intelligent entity – e.g., an AI agent, AAAI, or PSI – it is possible to have intelligent sub-components or sub-agents, as in mixture of experts approaches used in some LLMs. In this case, the universal problem solving architecture and all associated methods can also be used to coordinate intelligence “within the brain” so to speak of an individual intelligent entity like an AAAI. While the applicant has explained this inventive use of the systems and methods disclosed in this and other cited PPAs and PCTs, because this application is primarily focused on PI, the examples, system architecture, and disclosure primarily concern network of intelligent entities and the methods are described in the context of facilitating the collective cognitive abilities of such networks, and networks of networks, resulting in PI.

[00478] With respect to the architecture of PI, FIG. 109 simplifies this complexity by categorizing all the disclosed systems and methods as either: **10**, representing modular components of a Global Super Intelligent AGI Network (PI) or, **20**, as supportive systems and methods that can increase the safety, efficiency, and effectiveness of the PI system in various respects.

3.2 Major Components and Categories of Supporting Systems and Methods for Exemplary PI Architecture

[00479] There are many potential implementations of a PI architecture that are possible using various combinations of the systems and methods disclosed above and in previously cited PPAs and PCTs. There are also multiple ways to structure the major modular components referred to by **10** in FIG. 109, and the supportive systems and methods referred to by **20** in FIG. 109. For exemplary purposes, one exemplary implementation of an Architecture for Planetary Intelligence is illustrated in FIG. 110.

[00480] With reference to FIG. 110, **20** refers to the overall PI system. **21** refers systems and methods related to the online advertising technology present technologies that enable better monetization of online ads, and which can serve as source of funding and financial resources for the system. **22** refers to the universal problem solving framework, which serves as the means of

- rigorous communication for problem solving and other cognition across intelligent entities. **23** refers to a network that pools or coordinates multiple intelligent entities using the universal problem solving framework. **24** refers to the intelligence spot market, also referred to as “attention spot market” or “expertise spot market” or just “Spot Market” in previously cited PPAs and PCTs. **25** refers to Advanced Autonomous AIs, AAAs, sometimes referred to as AI agents, which can be customized and personalized. **26** refers to personalized AAAs that have sufficient capability and/or customization to exhibit super intelligent (i.e., intelligence surpassing the average human) behavior in certain areas, and which are known as Personalized SuperIntelligences, or PSIs, and which can work together on a network. **27** refers to the Artificial General Intelligence that results from the combined intelligence of multiple intelligent entities, which can include AAAs, AI agents, PSIs, and humans all working together. **28** refers to a network of AGIs, that incorporate a sense of awareness, and self-awareness, and multiple identities as described in previously cited PPAs and PCTs, and which together comprise a global superintelligent AGI network, or Planetary Intelligence.
- [00481]** The arrow from Online Ad Tech (**21**) to Universal Problem Solving Framework (**22**) represents the infusion of financial resources that can be associated with goals and problems on the CI network (**23**), and which can be used to reward problem solvers. Of course, financial resources can come from many sources, including clients will to pay for solutions to problems, but in this exemplary implementation, where methods for self-extending the network (**31**) apply, FIG. 110 is illustrating how the system might use money generated by the Online Ad Tech present technology to self-fund expansion of the network of AGIs and increase the scope of PI -- e.g. by using funds to reward solving the self-extending problem and/or recruiting intelligent entities via the Intelligence Spot Market (**24**).
- [00482]** The “+” between (**22**) and (**23**), and between (**23**) and (**24**) indicates that modules **22**, **23**, and **24** together comprise important elements of the network that work together in combination. This combination can be used internally:
- to by AAAs (as shown by the arrow to **25**), or
 - across multiple individual AAAs and humans or other intelligent entities in a network of AAAs (as shown by the two arrows to **26**), or
 - across networked PSIs, with other intelligent entities, to form AGI (as shown by the two arrows to **27**), or
 - across multiple AGIs, which themselves are networks of intelligent entities, to create a network of AGIs, which, if of sufficient global scope and capability, comprises a PI (as shown by the two arrows to **28**)

[00483] Further, the overall PI (20) will be a self-aware PI as long as its components have awareness, self-awareness, and identity-related capabilities enabled – e.g., by the systems and methods for these capabilities that have been previously disclosed.

5 [00484] The supportive systems and methods, generally referred to by 30, comprise four categories of inventive systems and methods that increase the profitability and scope (31), ethics and safety (32), speed of intelligence growth (33), and learning and knowledge combination abilities (34) of the overall PI system (20).

3.3 Detailed Mapping of Inventive Systems and Methods to Exemplary PI Architecture

10 [00485] This Section provides a detailed mapping of inventive systems and methods to the exemplary PI architecture disclosed in Section 3.2. A concise way of accomplishing this mapping, is to systematically examine each of the major components and categories of supporting systems and methods that were illustrated in FIG. 110, and list the specific inventive methods (by citing FIGS from Section 2.3) that can would be used, in an exemplary implementation, to implement the
15 function of each major component or category. Multiple methods can be used individually or in combination with other methods for each component or category. Therefore, in the lists that follow, it is understood that the applicant intends that “any one or any combination of” methods in the list can apply for implementation purposes.

[00486] With respect to FIG. 110, 21, the Online Ad Tech component, the relevant methods include,
20 without limitation, those described in FIGS. 76, 80, and 81.

[00487] With respect to FIG. 110, 22, the Universal Problem Solving Framework component, the relevant methods include, without limitation, those described in FIGS. 3, 6, 11, 12, 14, 15, 16, 20, and 21.

[00488] With respect to FIG. 110, 23, the Collective Intelligence Network component, the relevant
25 methods include, without limitation, those described in FIGS. 2, 4, 5 (items a-c, y, z, and a1 – m1), 7, 11 and 12 (the AAAI.com network items in these FIGS.), and the problem solving tree of FIG. 17.

[00489] With respect to FIG. 110, 24, the Intelligence Spot Market component, the relevant methods include, without limitation, those described in FIGS. 29, 46, 77, 78, 79, and 82.

30 [00490] With respect to FIG. 110, 25, the AAAIs component, the relevant methods include, without limitation, those described in FIGS. 1, 5, 10, 13, 26, 27, 30, 31, 36, and 47.

[00491] With respect to FIG. 110, 26, the Networked PSIs component, the relevant methods include, without limitation, those described in FIGS. 39, 40, 42, 44, 45, 74, and 75.

- [00492] With respect to FIG. 110, **27**, the AGI component, in addition to all the foundational methods enabling AGI, relevant methods include, without limitation, those described in FIGS. 28, 34, 35, 93, 94, 95, 96, 100, and 103.
- [00493] With respect to FIG. 110, **28**, the Self-Aware Networked AGIs component, the relevant methods include, without limitation, those described in FIGS. 83, 84, 85, 86, 87, 88, and 89.
- [00494] With respect to FIG. 110, **31**, the Self Extending & Profitable supporting category of methods, the relevant methods include, without limitation, those described in FIGS. 107 and 108 and those already mentioned as relevant to **20**.
- [00495] With respect to FIG. 110, **32**, the Safety and Ethics Checks / Designed Features supporting category of methods, the relevant methods include, without limitation, those described in FIGS. 18, 19, 23, 41, 43, 59, 60, 61, 62, 63, 64, 65, 66, 67, 69, 70, 71, 72, 73, 101, 102, and 104.
- [00496] With respect to FIG. 110, **33**, the Catalysts for Growth of Intelligence supporting category of methods, the relevant methods include, without limitation, those described in FIGS. 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, and 58.
- [00497] With respect to FIG. 110, **34**, the Learning and Knowledge Modules / Combination supporting category of methods, the relevant methods include, without limitation, those described in FIGS. 8, 9, 22, 24, 25, 32, 37, 38, 68, 90, 91, 92, 97, 98, and 99.

3.4 Exemplary Specific PI Implementation Using Subset of Systems and Methods

- [00498] Imagine that you are the CEO of a major technology company, such as META, and you wish to implement a Planetary Intelligence system comprised of a network of powerful AGI systems. You might proceed as follows.
- [00499] 1. With reference to FIG. 110 (**31**): You want the PI system to be self-funding and self-extending. Since META is already in the online advertising business, you decide to implement the online ad technology technologies disclosed in PPA/PCT#8. You already have advertising technology systems and infrastructure similar to that described in FIG. 76.
- [00500] You instruct your development team to modify the “creative” of a large number of ad units so that they can work with the online problem solving system described in this and cited PPAs and PCTs.
- [00501] FIG. 80 describes some of the modifications that the team needs to make. You monetize the expertise of human experts targeted by the modified ad technology by selling the expertise and data gathered to other companies that are desperately seeking new sources of training data and expertise for their AI efforts. You can charge higher rates than a normal ad would generate for this type of valuable information that your modified ad tech is scooping up. You also use the data from human

experts to train your own AI models, and the human experts themselves will help power the PI system.

[00502] 2. With reference to FIG. 110 (22): You need to funnel the human experts gathered in Step 1. to an online problem solving system, where they will use the Universal Problem Solving Framework to collaborate with other intelligent entities – initially mainly other humans. FIG. 6 describes the Universal Framework the human experts will use.

[00503] FIGS. 11 and 12 describe at a high level how they will solve problems, sequentially or in parallel.

[00504] FIG. 14 describes some of the methods, including matching human experts to problems, compensating the solvers, etc. that will be needed to implement the problem solving system.

[00505] FIG. 16 identifies some key characteristics of problem solving, such as a goal sub-goal hierarchy and means for recording solution attempts in an auditable record that will facilitate learning by the system.

[00506] You do not want human experts to have to know anything about problem-solving theory or the universal architecture, so the system uses a Natural Language to Problem Solving Language Translator (as described in FIG. 21) that enables the humans to work with regular natural language while the system translates everything into the rigorous problem solving framework that enables collaboration with other (human and non-human) intelligent entities.

[00507] 3. With reference to FIG. 110 (23) problem solving takes place on a network, as described in FIG. 2.

[00508] Further operation of this network is described in detail in FIG. 5 (items **a-c**, **y**, **z**, and **a1-m1**).

[00509] The network tracks all problem solving activity in a giant problem solving tree structure, a.k.a. the “WorldThink Tree”, which is described in FIG. 17.

[00510] 4. With reference to FIG. 110 (24), for specific problems, META may purchase expertise, human attention, or attention from other intelligent entities via an intelligence spot market. META is an excellent position to actually run the spot market, since it has access to so many human users that might wish to participate. Since the value of human attention is linked to reputation metrics, META will likely administer an impartial reputational system as described in FIG. 29, to help categorize the types of expertise offered on the spot market and facilitate price discovery. This same reputational system could also be used by the collective intelligence network in step 3, and reputations are also associated with AAAs discussed below in step 5.

[00511] Standardized tasks, as described in FIG. 46, are important ways of benchmarking the performance of (human or non-human) intelligent entities, which can also affect pricing on the spot market.

5 [00512] FIGS. 77, 78, and 79 provide additional details on the methods for implementing the spot market itself.

[00513] Finally, FIG. 82 describes how to optimize parameters relating to the operation of the spot market. Since attention, expertise, and knowledge of intelligent entities is likely to become the most valuable commodity in the future, operating a spot market for them is both a lucrative source of revenue to fund PI, and also a means of ensuring that the PI has access to the best intelligence
10 available for any given task.

[00514] 5. With reference to FIG. 110 (25) META may choose to augment human experts on the networks with AI agents. Fortunately, META has developed some foundational AI agents itself (e.g. Llama 3) and has access to many more. However, in order to be most effective as intelligent entities on the network, these foundational model have to be trained, tuned, and customized further. They
15 need to become Advanced Autonomous AIs, or AAAIs. The roles that these AAAIs will play, how they can be customized, how they can use a problem solving architecture, how they fit on the network, how they integrate, and how they improve is generally outlined in FIG. 1.

[00515] A detailed implementation process for creating, training, and increasing the intelligence of these AAAIs is shown in FIG. 5 in the “AAAI creation” section, while the top half of FIG. 5 shows
20 how they participate on the network to solve problems.

[00516] Another view of how they integrate, including with safety functions, is described in FIG. 10.

[00517] To enhance their effectiveness, once trained, the AAAIs can be cloned as shown in FIG. 13. Cloning allows many cloned AAAIs to participate in multiple problems in parallel.

25 [00518] Because META has access to so many human users, and has extensive social media profile information on billions of users, this information can be used by META – or by META’s human users – to customize and train customized versions of AAAIs, as described in FIG. 26.

[00519] An important aspect of this customization and training is that as each AAAI is trained, it learns not only knowledge and expertise from a human user (and the user’s profile) but also ethical
30 and values information. The result is that META ends up owning, or having access to, billions of customized AAAIs, each reflecting slightly different values and ethical preferences. When aggregated, the preferences of the AAAIs constitute a representative and statistically valid sample of human values and ethics, which is essential for creating safe and ethical AGI.

[00520] Various methods, as illustrated in FIG. 27 can be used to train and customize the AAAIs.

[00521] META may choose to use additional methods related to the development and deployment of AAAs as described in FIGS. 30 and 31.

[00522] To help prevent the AAAs from hallucinating, when collaborating on problem solving, and to increase their reliability, META may want to implement the method described by FIG. 36.

5 [00523] Finally, to optimize the performance of AAAs that have been customized as PSIs, META can use the methods of FIG. 47, producing multiple versions and optimizing parameters for problem solving.

[00524] 6. With reference to FIG. 110 (26), META will want to upgrade its AAAs into PSIs by customizing personalizing them, e.g., by using methods described in FIG. 39.

10 [00525] PSIs will improve further as they work over long periods in a variety of scenarios (FIG. 40), record and learn from solution attempts (FIG. 42) and evolve via competition and selection (FIG. 44).

[00526] The PSIs will work on a network as discussed in FIG. 45, where among other things they can combine ethical preferences (FIG. 74) and collectively, with other intelligent entities, comprise and AGI (FIG. 75).

[00527] 7. With reference to FIG. 110 (27), FIG. 28 describes generally how scalable and ethical AGI can arise from multiple intelligent entities that engage in problem solving and refine their values.

[00528] FIG. 34 shows that one method for combining information, including ethical information, from multiple base LLMs or AAAs or PSIs is to experiment with combinations of the weight matrices in which these entities store their knowledge.

[00529] FIG. 35 expands upon this method.

[00530] FIGS. 93, 94, 95, and 96 describe various methods for implementing identities that can aid in formulating a sense of self-awareness for these entities, and also for resolving potential conflicts resulting from different identities. The process of identity formation and conflict resolution will aid the AGI in developing a robust sense of self-awareness.

[00531] Further methods for dealing with identity conflicts, including methods of negotiation and compromise are described in FIGS. 100 and 103. For safety and ethical reasons, META would likely want to implement these and other related methods before enabling full self-awareness at the AGI level.

[00532] 8. With reference to FIG. 110 (28), a PI should ultimately have a sense of global awareness, including mechanisms for tracking current awareness, and locating its own sense of self-awareness within its current awareness as illustrated in FIG. 83.

- [00533] The PI (an even AGI, PSI and AAAI systems) must be able to maintain multiple identities simultaneously and understand the relationships between these identities as illustrated in FIG. 84. While awareness and self-awareness could be implemented at the AI, AAAI, PSI, or AGI level, in this example we are assuming that PI will be self-aware and networked.
- 5 [00534] To implement self-awareness, the PI must be equipped with attentional components as described in FIG. 86.
- [00535] Working memory parameters can adjust the scope of attention as described in FIG. 87.
- [00536] The PI must be capable of modelling awareness as described in FIG. 85, and monitoring and updating its awareness as described in FIG. 88.
- 10 [00537] For the awareness system to be efficient and safe, there must be attentional interrupts as described in FIG. 89.
- [00538] To reiterate (and as illustrated in the middle two boxes of FIG. 111), AGI arises out of the combined collective intelligence of individual intelligent entities working together on a problem solving network. PI arises out of the combined collective intelligence of AGIs, that have been
- 15 networked together to extend the scope and capabilities of the PI system.
- [00539] The PI becomes self-aware, if methods for enabling attention, awareness, self-awareness, identity, and identity conflict resolution (as described in Step 8) are activated. While awareness can be activated at levels below PI, such awareness is likely essential at the PI level in order for the system to function effectively given its global scope.
- 20 [00540] The steps 1-8 above, provide a skeletal implementation example of PI, but many other methods are desirable, and in some cases essential for effective and efficient operation. With reference to FIG. 110, these addition methods have been categorized and supportive methods, referred to generally by numeral (30). We now discuss some of these supportive methods and how they enhance the skeletal PI system just described.
- 25 [00541] With reference to FIG. 110 (31), a general method for automatically extending the network of AGIs and thus increasing the scope of the PI would be extremely helpful and has been described in FIGS. 107 and 108.
- [00542] Further PPA/PCT #8 describes novel online ad technology and novel technology to enable an intelligence, attention, or expertise spot-market. These present technologies can generate
- 30 significant financial resource which are necessary to help the PI self-extend. One can imagine other problem solving tasks – e.g., “develop a quantitative method for generating reliable and attractive returns with minimal risk in the US equity markets” – that could be posed to AAAI, PSI, AGI or PI systems to generate further funding for expansion. Ultimately, there are direct correlations between the amount of intelligence that can be focused on a problem and the amount of profits that can be

generated, provided profitable problems are chosen. Given the projected superior intelligence of AGI and PI systems to any one or group of humans, it is reasonable to expect that such systems will ultimately self-fund their growth in intelligence.

- [00543]** Because PI is so powerful, the cited PPAs and PCTs disclose dozens of safety mechanisms, methods, and features to align PI with human values and ensure human safety. For example, with reference to FIG. 110 (**32**); FIGS. 18 and 19 describe one of the most important and most scalable safety present technologies, namely embedding of safety checks in the universal problem solving framework itself, such that no goal or sub-goal can be set, and no problem solving action taken, unless the goal or action passes a safety and ethics check.
- 5
- [00544]** This checking process, because it is embedded in the operation of the problem solving process, scales with the speed of problem solving. That is, if PI can take one billion problem solving steps in a millisecond, it must also perform one billion safety checks. Although human cognition cannot keep track of problem solving at this speed, the safety checks are still being run, extremely quickly.
- 10
- [00545]** Further, there are ethical dimensions to these checks, which might be elicited as described in FIG. 23.
- 15

[00546] One of the most powerful ideas for keeping PI safe, even after it greatly outstrips human cognitive abilities, is the community safety method described in FIG. 41.

- [00547]** In this method, which applies to AAAs, PSIs, AGIs, and PI, humans rely on a community of entities that are much smarter and more powerful than humans. Each entity was trained and customized initially on human values. While it is possible, and perhaps likely, that a few of these Super Intelligent entities may evolve and change their values such that they become malevolent towards humans, it is far less likely that the majority of intelligent entities evolve in this way. The method of FIG. 41 halts activities of the dangerous entities via oversight by human-aligned entities.
- 20
- As long as a majority of the intelligence remains human-aligned, this mechanism for advanced AI policing advanced AI should keep humans safe in the same way that Bitcoin is un-hackable as long as the majority of participants in the system are honest.
- 25

[00548] FIG. 42 describes how all problem solving activities is recorded, e.g., using auditable and transparent blockchain technology.

- 30 **[00549]** This record enables safety checks as described in FIG. 43.

[00550] Simulations, with human validation, offer another avenue to increase safety as described in FIG. 59.

[00551] With regard to ensuring ethical behavior of systems, multiple philosophically-based approaches are possible, including but limited to the Consequentialist Approach (FIG. 69), the

Deontological Approach (FIG. 70), the Virtue Ethics Approach (FIG. 71), and the Golden Mean Method (FIG. 72).

[00552] The methods of FIG. 73 can be used to help ensure compliance of the AGI or PI with local and global regulations and ethical norms.

- 5 [00553] In situations where certain identities may lead to unsafe or unethical actions on the part of the AGI or PI, methods that allow human override – such as that described in FIG. 101 – can be used.

[00554] In less threatening cases of internal ethical conflict, arbitration with input from humans or other intelligent entities can help an AGI or PI resolve issues, as described in FIG. 102.

- 10 [00555] The ability to halt actions, including suspending identities that lead to unsafe or unethical actions by an AGI or PI, are critical safeguards, as described, for example, in FIG. 104.

[00556] Many safety and ethics safeguards rely on the fact that the very architecture of the PI is composed on personalized and customized intelligent entities, each of which reflect values of humans, and, which, in aggregate form a representative and valid sample of human values. That is,
15 the design of the AGI and PI networks is democratic in nature and thus offers the same level of ethics that humans are familiar with in democratic societies.

[00557] FIGS. 60 and 61, describe standard and weighted voting methods that help an AGI or PI integrate values democratically.

- 20 [00558] Methods for delegating voting authority, as described in FIG. 65, are also helpful in this regard.

[00559] FIG. 66 describes an innovative approach to using reputational concepts to preserve minority viewpoints and avoid “herd” decisions that might result in unethical or unsafe decisions by an AGI or PI.

- 25 [00560] AGI and PI systems can use methods based on converging evidence, as described in FIG. 64, to help identify which human values apply in a given situation.

[00561] Experiments, focus groups, interviews with humans, and other methods as described in FIG. 63 can also be helpful.

- 30 [00562] Democratic voting methods to help AGI and PI systems adopt or combine value systems are described extensively in a range of methods such as is imperfect, but it avoids concentration of power, which history has shown to be a major threat to human well-being in cases where the values of the single powerful entity are not aligned with the welfare of the people.

[00563] With reference to FIG. 110 (33), once META (in our example), develops a large number of AAAs which are customized and personalized, the capabilities of META’s AGI network depends critically on the abilities and intelligence of these component AAAs. Ideally, they are PSIs –

capable of SuperIntelligent performance in at least some areas of cognitive activity. Many PSI would enable a SuperIntelligent AGI. However, regardless of the level of intelligence exhibited by META's AGI or PI network, competitive advantage comes from increasing the intelligence of these entities as quickly as possible. The methods for catalyzing the growth of intelligence are relevant
5 here.

[00564] For example, across all of the component PSIs, and for the AGI and PI networks, the most useful new sources of information for increasing that intelligence should be identified as described in FIG. 49.

[00565] Every source of potential new information should be categorized along multiple dimensions
10 as described in FIG. 50.

[00566] Algorithms related to the compressibility of the information could be applied to determine the information content of the new information sources, as described in FIG. 51.

[00567] Parameters from Kaplan Information Theory (KIT) could be applied to further assess the usefulness of the information sources as described in FIG. 52.

[00568] The distribution of tasks (and associated goals) that META's AGI or PI encounters could be
15 calculated and then the method described in FIGS. 53 and 54 could be applied to help prioritize new dataset that might most quickly improve the capabilities of the AGI or PI relative to its most frequent goals.

[00569] The AGI or PI could balance the cost of acquiring information against its usefulness as
20 described in FIG. 55.

[00570] Different versions of the PI, AGI, or of individual PSIs could be compared in terms of their effectiveness, and methods described in FIGS. 56 and 57 could be further used to increase the intelligence of these entities.

[00571] Finally, heuristics described in FIG. 58, such as seeking information that is maximally
25 different from the entity's existing views, or seeking more recent information, or using principles such as converging evidence to increase reliability, can be employed.

[00572] With reference to FIG. 110 (34), PI must continually learn and acquire new knowledge in order to increase its cognitive capabilities. One way that an AGI or PI system can learn is by storing and indexing successful solutions to problems on the network, via procedural learning methods as
30 described in FIGS. 8 and 9.

[00573] Learning at the AAAI level also helps increase the cognitive capabilities of an AGI or PI by increasing the intelligence of the component entities. For example, FIG. 22 describes training methods that can help AI entities learn, become more customized and gain specialized knowledge.

[00574] Acquiring new data, often from humans, is often a prerequisite for further learning by an AGI or PI. In this regard, access to methods that extract such information, like the methods for automatic generation of questionnaires described in FIG. 24, or the methods for identifying patterns in human behavior described in FIG. 25, can be helpful.

- 5 [00575] AGI and PIs may be able to rapidly increase their knowledge by direct combination of knowledge that is contained in the weight matrices of their constituent AAAs, e.g., as described in FIGS. 32 and 37.

[00576] As described in FIG. 38, combinations of knowledge from multiple entities can be accomplished either by directly combining weight matrices, or by combining training datasets and then training constituent entities of an AGI or PI directly from the combined training data.

[00577] PIs can learn information related to safety and regulation compliance via multiple methods for training their constituent entities, some of which methods are described in FIG. 68.

- [00578] Since PIs will have many different identities, due to their vast scope, general methods for learning new identities, or changing an entity's sense of identity, can be helpful as described in FIG. 90.

[00579] META (in this example) would want its AGI or PI to attain a level of intelligence where it could interact with humans as capably as the most capable humans. Using methods such as the modified "Turing Test" described in FIG. 92, can help evaluate when the PI system has achieved the desired level of cognitive performance.

- 20 [00580] The more complex an AGI or PI becomes, the more important simulation and scenario modelling become. Therefore, methods such as those described in FIGS. 97 and 98 can be especially helpful in maximizing the chances that AGI or PI acts in positive, human-aligned ways.

[00581] Finally, if PI is to be safe for humanity, it is important for PI to collaborate and continuously co-evolve with human society, as described in the methods of FIG. 99.

- 25 [00582] If META's PI system incorporated all of the methods above, in an exemplary implementation, it would be much more capable, powerful, safe, and human-aligned than any AI system that currently exists. Other combinations of methods cited in PPA/PCTs #1 – 10 are possible and may be desirable depending on the specific goals and purposes of the PI. However, the above exemplary implementation represents a novel and useful initial design for PI.

- 30 [00583] FIG. 112 is a diagrammatic representation of a computer system **100** that is utilizable or implementable with the user's device and/or any peripheral component of the present technology. The computer system **100** can be part of an example machine, which is an example of one or more of the computers referred to herein and, within which a set of instructions for causing the machine to perform any one of or more of the methodologies discussed herein may be executed. In various

example embodiments, the machine operates as a standalone device or may be connected (e.g., networked) to other machines. In a networked deployment, the machine may operate in the capacity of a server or a client machine in a server-client network environment, or as a peer machine in a peer-to-peer (or distributed) network environment. The machine may be a personal computer (PC), a tablet PC, a set-top box (STB), a personal digital assistant (PDA), a cellular telephone, a web appliance, a network router, switch or bridge, or any machine capable of executing a set of instructions (sequential or otherwise) that specify actions to be taken by that machine.

5 [00584] The computerized system **100** can include one or more processors **102**, storage devices **106**, and communication devices, as well as software components or instructions **104** for providing a platform for users to interact with and train/tune the LLMs. The computing capabilities may be stand alone or may be cloud based. They may include cloud based AI development platforms that seamlessly offer “AI as a service” and they may include both hardware and software components.

10 [00585] The system also supports the ability for users to provide new data, or data that is unique to them, for the LLMs to learn from. The processors **102** may be one or more CPUs, GPUs, chips specialized for ML, microprocessors, application processors, embedded processors, field-programmable gate arrays (FPGAs), or other hardware components capable of executing computer programs. The processors may be in communication with one another and/or with other components of the system. Further, any one of or any combination of the components of the system **100** can communicate with each other via a bus **134**.

15 [00586] The storage devices **106** may include one or more hard drives, solid-state drives, optical storage devices, or other storage components. The storage devices may store the data that is used to train/tune the LLMs, as well as other data associated with the system, such as user accounts, system settings, and other data.

20 [00587] The communication devices may include one or more cellular modems **108**, Wi-Fi cards **110**, Bluetooth modules **112**, Network Interface Device **114**, or other components that enable the system to communicate with other systems, such as user devices, over a network or the internet.

25 [00588] The communication devices may also enable the system to communicate with other systems over a wireless or wired connection **116**.

30 [00589] The software components may include computer programs for providing a platform for users to interact with and train/tune the LLMs. The software components may also include computer programs for collecting, storing, and processing data that is used to train and/or tune the LLMs. The software components may also include computer programs for providing a user interface for users to interact with the system.

[00590] The user interface 118 may include, without limitation, natural language interfaces, textual interfaces, and chatbot type of interfaces, a web-based user interface, a mobile application, an augmented reality application, a metaverse application, or other applications that allow users to interact with the system. The user interface may include features for allowing users to select the data that they want to use to train/tune the LLMs, as well as features for allowing users to interact with and monitor the progress of the LLMs.

[00591] The system may also include one or more datacenters, databases or data sources, including without limitation vector databases, centralized databases, and distributed databases, for storing the data that is used to train/tune the LLMs, as well as other data associated with the system, such as user accounts, system settings, and other data. The databases may be hosted on the system itself or on another system, including cloud based systems.

[00592] The system may also include one or more authentication systems for verifying the identity of users who use the system, as well as for providing secure access to the system. The authentication systems may include biometric authentication systems 122, such as facial recognition or fingerprint recognition systems, as well as other authentication systems, such as password-based authentication systems.

[00593] The system may also include one or more security systems for protecting the system from unauthorized access and for protecting the data that is stored on the system. The security systems may include firewalls, encryption systems, access control systems, single and multi-factor authentication systems, and other security systems.

[00594] The system may also include one or more analytics systems for collecting and analyzing data associated with the system and/or the LLMs. The analytics systems may include machine learning algorithms and other algorithms for analyzing the data associated with the system and/or the LLMs.

[00595] Data visualization methods, including use of problem trees and other representations and data structures; use of statistical outputs, tables, graphs, text, speech, video, image and graphical outputs may be used for one way or di-directional communication between users and the system, and between multiple (human or AI) agents or LLMs using the system to interact with each other in large or small groups.

[00596] The system may also include one or more monitoring systems for monitoring the performance of the system and/or the LLMs. The monitoring systems may include systems for monitoring the performance of the system, such as system uptime, and systems for monitoring the performance of the LLMs, such as accuracy, speed, ethical compliance, reputation metrics, quality metrics, and other metrics as discussed above or as are known in the art.

- [00597] The system may include one of more of the architectures described above that enable one or more human or AI Agents or LLMs to engage in a variety of intellectual tasks including, without limitation, simple and complex and multi-step problem solving behavior with the system having all of the functionality and features previously described.
- 5 [00598] The system may also include one or more feedback systems for allowing users to provide feedback on the system and/or the LLMs. The feedback systems may include systems for allowing users to submit feedback on the system, such as bug reports, and systems for allowing users to submit feedback on the LLMs, such as suggestions for improving the accuracy or speed of the model.
- 10 [00599] The system may also include one or more management systems for managing the system and/or the LLMs. The management systems may include systems for managing the system, such as systems for managing the users and user accounts, and systems for managing the LLMs, such as systems for managing the data used to train and/or tune the model.
- [00600] The system may also include one or more payment systems allowing users to pay for the use of the system and/or the LLMs. The payment systems may include systems for processing payments, such as credit card processing systems, and systems for managing payments, such as subscription management systems.
- 15 [00601] The system may also include one or more other components, such as support systems, reporting systems, and other components that are necessary for providing a platform for users to interact with and train/tune the LLMs.
- 20 [00602] The computerized system of the present technology enables users to interact with and train/tune LLMs based on data that is unique to the users. The components of the system described herein provide the necessary hardware and software components for enabling users to do so.
- [00603] Further, while only a single machine is illustrated, the term “machine” shall also be taken to include any collection of machines that individually or jointly execute a set (or multiple sets) of instructions to perform any one of or more of the methodologies discussed herein.
- 25 [00604] The computer system **100** may further include or be in operable communication with a video display **120** (e.g., a liquid crystal display (LCD), touch sensitive display), input and/or output device(s) **130** (e.g., a keyboard, keypad, touchpad, touch display, buttons, sonic, sensorial, etc.), a cursor control device **132** (e.g., a mouse), a drive unit **124** (also referred to as disk drive unit), and a signal generation device **128** (e.g., a speaker). The drive unit **124** can include a computer or machine-readable medium **126** on which is stored one or more sets of instructions and data structures (e.g., instructions **104**) embodying or utilizing any one of or more of the methodologies or functions described herein. The instructions **104** may also reside, completely or at least partially,
- 30

within the memory **106** and/or within the processors **102** during execution thereof by the computer system **100**. The memory **106** and/or the processors **102** may also constitute machine-readable media.

5 **[00605]** Still further, the computer system **100** can be in operable association or communication with any type of multi-modal input and/or output **130** that address the human senses, as well as I/O technology that extends beyond the range of normal human perception. Such as the ability to process invisible to humans, for example but not limited to, Xrays and information outside of the typical bandwidths of human perception, but not outside of AI perception using tools. Additionally, the I/O technology can include very fast perceptions that are too fast for humans to perceive but
10 which an AI entity could perceive, and very slow or faint perceptions (e.g., tiny seismic shifts occurring over years) that humans cannot perceive but which AIs could. Since any intelligent entity can be part of the present technology system, then it can be appreciated that any type of I/O that humans, and also AIs with much broader perceptual capabilities than humans, can be utilized with the system **100**.

15 **[00606]** The instructions **104** may further be transmitted or received over a network via the network interface device **114** utilizing any one of a number of well-known transfer protocols (e.g., Hyper Text Transfer Protocol (HTTP)). While the machine-readable medium is shown in an example embodiment to be a single medium, the term “computer-readable medium” should be taken to include a single medium or multiple media (e.g., a centralized or distributed database, vector
20 databases, and/or associated caches and servers) that store the one or more sets of instructions. The term “computer-readable medium” shall also be taken to include any medium that is capable of storing, encoding, or carrying a set of instructions for execution by the machine and that causes the machine to perform any one of or more of the methodologies of the present application, or that is capable of storing, encoding, or carrying data structures utilized by or associated with such a set of
25 instructions. The term “computer-readable medium” shall accordingly be taken to include, but not be limited to, solid-state memories, optical and magnetic media, and carrier wave signals. Such media may also include, without limitation, hard disks, floppy disks, flash memory cards, digital video disks, random access memory (RAM), read only memory (ROM), and the like. The example embodiments described herein may be implemented in an operating environment comprising
30 software installed on a computer, in hardware, or in a combination of software and hardware.

[00607] An example machine system of the present technology including the computer system **100** in combinational and/or operational use with components of the present technology. In the exemplary, any or all of above described components can include a processor **102**, memory **106**, a network interface device **114**, a display **120**, an input device(s) **130**, **132**, and/or drive unit **124**.

[00608] According to one aspect, the present technology can include a system for PI with human-aligned behavior utilizing a collective network of intelligent entities selected from the group consisting of any combination of a user computer system operated by a human user, an AI agent or system, an AAAI agent or system, an AGI agent or system, a SI agent or system and a PSI agent or system. The system can include multiple intelligent entities each connected to a collective network and each comprising: a processor, a computer-readable storage medium, and program instructions stored on the computer-readable storage medium being executable by the processor to cause the intelligent entities to each independently:

- utilize a modular architecture configured or configurable to scale from components within an individual intelligent entity on the collective network;
- implement a universal problem solving architecture and framework on a task to collaborate and create higher levels of intelligence;
- utilize a human-centered input configured or configurable for obtaining values and ethics information from multiple human users and then using the values and ethics information to customize the AI agent or system;
- combine data from multiple of the intelligent entities at a level of the AAAI;
- customize the PSI agent or system using intelligent agents configured or configurable to develop and continuously improve the PSI agent or system;
- increase a knowledge of the intelligent entities by incorporating new sources of informational datasets that differs from a current informational dataset of the intelligent entity;
- combine the values and ethical information of other intelligent entities, and to resolve conflicts between different value systems;
- utilize online advertising technology for increasing an intelligence of the intelligent entities;
- create a self-aware operation for the AGI or PI agent or system by adding a dimension of self-awareness and increased autonomy to the AGI or PI agent or system, and to create an ability to assume multiple identities of the AGI or PI agent or system to handle tasks that arise in parallel with other AGI or PI agents or systems; and
- search through any one of or any combination of a collective network of AIs, a collective network of AAAIs, a collective network of AGIs, and a collective network of PIs for new information that is different to a current information, respectively, and to incorporate the new information to the current information, respectively.

[00609] According to another aspect, the present technology can include a method for PI with human-aligned behavior utilizing a collective network of intelligent entities selected from the group consisting of any combination of a user computer system operated by a human user, an AI agent or

system, an AAAI agent or system, an AGI agent or system, a SI agent or system and a PSI agent or system. The method can include the steps of:

- utilizing a modular architecture configured or configurable to scale from components within an individual intelligent entity on the collective network;
- 5 implementing a universal problem solving architecture and framework on a task to collaborate and create higher levels of intelligence;
- utilizing a human-centered input configured or configurable for obtaining values and ethics information from multiple human users and then using the values and ethics information to customize the AI agent or system;
- 10 combining data from multiple of the intelligent entities at a level of the AAAI agent or system; customizing the PSI agent or system using intelligent agents configured or configurable to develop and continuously improve the PSI agent or system;
- increasing a knowledge of the intelligent entities by incorporating new sources of informational datasets that differs from a current informational dataset of the intelligent entity;
- 15 combining the values and ethical information of other intelligent entities, and resolving conflicts between different value systems;
- utilizing online advertising technology for increasing an intelligence of the intelligent entities;
- creating a self-aware operation for the AGI or PI agent or system by adding a dimension of self-awareness and increased autonomy to the AGI or PI agent or system, and creating an ability
- 20 to assume multiple identities of the AGI or PI agent or system to handle tasks that arise in parallel with other AGI or PI agents or systems;
- creating the PI agent or system comprising a collective network of AGIs, wherein the collective network of AGIs includes any one of or any combination of a collective network of the user computer systems, a collective network of AIs, a collective network of AAAs, and a
- 25 collective network of PSIs; and
- searching through any one of or any combination of the collective network of the user computers systems, the collective network of AIs, the collective network of AAAs, the collective network of AGIs, the collective network of PSIs, and a collective network of PIs for new information that is different to a current information, respectively, and incorporating
- 30 the new information to the current information, respectively.

[00610] In some embodiments, the step of utilizing the modular architecture can include the steps of:

- customizing one or more attributes of the intelligent entity;

integrating one or more datasets from any one of or any combination of the intelligent entities;
and

improving, by utilizing one or more techniques, any one of or any combination of the
customizing of the attributes, the universal problem solving architecture and framework, the
collective network and the integrating of the datasets.

[00611] Some embodiments of the present technology can include a step of assigning a credit value
or a blame value to the datasets based on whether the datasets increase or decrease performance of
the intelligent entities based on performance metrics or evaluation functions.

[00612] Some embodiments of the present technology can include a step of quantifying a benefit
weight or a harm weight to a contribution by each of the intelligent entities to the task.

[00613] Some embodiments of the present technology can include a step of distributing a reward to
the intelligent entities proportionally to the contribution of the intelligent entities based on the
benefit weight or the harm weight.

[00614] In some embodiments, the step of implementing the universal problem solving architecture
and framework can include the steps of:

acquiring information associated with the task from any one of or any combination of the
intelligent entities;

identifying one or more of the intelligent entities that have one or more attributes related to one
or more request criteria of the task;

implementing by each of the identified intelligent entities the universal problem solving
architecture and framework on the task to create a completion solution; and

providing the completion solution to any one of or any combination of the intelligent entities for
final acceptance.

[00615] Some embodiments of the present technology can include the steps of:

executing an ethics check on the task, and a solution for the task provided by any one of or any
combination of the intelligent entities;

comparing any one of or any combination of the task, and the solution against prohibited
attributes, and assigning an ethics attribute to one of or any combination of the task, and the
solution based on any one of or any combination of a result of the comparison, and an ethics
criteria;

implementing, based on the result of the comparison, the universal problem solving architecture
and framework on the task to create the solution and creating an AGI; and

providing the results of the comparison and the solution to any one of the intelligent entities and
additional intelligent entities on the collective network.

[00616] Some embodiments of the present technology can include a step of recording one or more problem solving activities from each of the intelligent entities in an auditable record, and comparing the problem solving activities with a successful or unsuccessful progress towards the solution of the task, and determining which of the problem solving activities to keep active.

- 5 [00617] Some embodiments of the present technology can include a step of learning by the intelligent entities a procedural learning process of the universal problem solving architecture and framework, wherein the intelligent entities provide information to the procedural learning process for creation of the AGI.

- 10 [00618] Some embodiments of the present technology can include a step of cloning any one of the intelligent entities for deployment of multiple copies thereof to assist in any one of or any combination of creating of the solution, or providing training data to any one of the intelligent entities.

- 15 [00619] Some embodiments of the present technology can include a step of estimating a worth of the cloned intelligent entities utilizing a network effect value including the number of cloned intelligent entities s available on the collective network.

[00620] Some embodiments of the present technology can include a step of utilizing the estimated worth for determining pricing decisions for problem solving services offered by the cloned intelligent entities on any one of the social media platforms or through an additional intelligent entity.

- 20 [00621] Some embodiments of the present technology can include a step of monetizing the cloned intelligent entities for each utilization of the cloned intelligent entities on the social media platforms or the additional intelligent entity.

- 25 [00622] Some embodiments of the present technology can include a step of allowing access to the cloned intelligent entities by any one of the social media platforms so that the social media platforms can receive the solution to the task or using the training data for an AI system of the social media platform.

[00623] In some embodiments, the step of utilizing the human-centered input can include the steps of:

- 30 matching one or more human workers from a data source including a list of human problem solvers to the task based on a task criteria;
translating any part of the task into an unambiguous language utilizable in the universal problem solving architecture and framework including a decision tree;
separating the task into sub-tasks;

delegating each of the sub-tasks to one or more of the matched human workers so that work on each of the sub- tasks proceeds independently from each other and parallel with each other; utilizing the universal problem solving architecture and framework in a problem solving process on the sub-tasks, respectively, to create one or more sub-solutions;

5 receiving the sub-solutions from each of the matched human workers for the sub-tasks delegated thereto;

combining the sub-solutions into an overall solution to the task;

directing any one of or any combination of a new human worker from the data source and one or more of the matched human workers to parts of the decision tree where work is required;

10 compensating the matched human workers for the sub-solutions;

providing any one of or any combination of the sub-solutions and the overall solution to any one of or any combination of the intelligent entities;

allowing any one of or any combination of the intelligent entities to accept the overall solution, reject the overall solution, and provide feedback to any one of the matched human workers

15 on any one of the sub-solutions; and

assigning a reputation attribute to any one of or any combination of the human workers and the intelligent entities.

[00624] In some embodiments, the decision tree is maintained in blockchain or Ethereum logs.

[00625] In some embodiments, the reputation attribute includes metrics on any one of or any combination of a time to the sub-solutions, a difficulty value of the task, short and long-term user satisfaction with the sub-solutions, a number of times any one of the sub-solutions has been re-used on the collective network, a rating by other human workers, a responsiveness value of the human workers, and a reliability value of the human workers.

20

[00626] Some embodiments of the present technology can include a step of using the reputation attribute in the matching of the human workers to the task using an algorithm to the delegation of the sub-task.

25

[00627] Some embodiments of the present technology can include a step of soliciting, at predetermined intervals after the overall solution or the sub-solutions are provided to any one of the intelligent entities, feedback by way of a survey for user satisfaction information to obtain satisfaction metrics that are used to update the reputation attribute of one or more of the human workers or the intelligent entities, respectively.

30

[00628] In some embodiments, the step of combining data from multiple of the intelligent entities at a level of the AAAI agent or system can include the steps of

training a base Large Language Model (LLM) of a first AI agent or system with guardrails including attributes associated with any one of or any combination safety, ethics and knowledge;

customizing the base LLM to an ethics profile associated with a first human user;

5 combining ethical information from multiple intelligent entities different to that of the first AI agent or system and the first human user;

refining a set of values of the base LLM based on problem solving on the task; and

updating the base LLM with the combined ethical information and the refined set of values thereby allowing for a scalable AGI.

10 **[00629]** In some embodiments, the customizing of the base LLM can include the step of assembling a corpus of ethical questions based on various ethical assessment instruments and supplemented by first questions based on data from a social media platform and second questions solicited from crowdsourcing.

[00630] Some embodiments of the present technology can include a step of assigning regression weight values to the ethical questions such that a ranking is achieved whereby higher-ranked questions provide more useful ethical information than lower-ranked questions.

[00631] Some embodiments of the present technology can include a step of combining weight values from the intelligent entities with the regression weight values for improving a tuning of any one of the intelligent entities.

20 **[00632]** In some embodiments, the step of customizing the PSI agent or system using intelligent agents configured or configurable to develop and continuously improve the PSI agent or system can include the steps of :

acquiring a base-level AI agent or system that has previously been customized;

collecting media information related to the human user of the base-level AI agent or system;

25 analyzing the media information;

transforming the analyzed media information into training data sets;

differentially weighting the transformed training data sets;

adding knowledge modules to the weighted transformed training data sets;

locating new sources of data to include to the weighted transformed training data sets;

30 applying the weighted transformed training data sets to the base-level AI agent to create a user PSI; and

communicating the user PSI with multiple additional PSIs using the collective network to enable community-based safety features from the additional PSIs to the user PSI.

[00633] Some embodiments of the present technology can include the steps of:

communicating the user PSI with multiple additional intelligent entities using the collective network to enable community-based safety features, wherein the user PSI and the additional intelligent entities each agree to use a set of rules relating to safety or ethics;

recording all actions by the user PSI and the additional intelligent entities in an auditable form

5 on any one of or any combination of a central computer system on the collective network, the intelligent entities, and the additional intelligent entities; and

monitoring that each of the actions follow the set of rules, and flagging any of the actions that do not follow the set of rules.

[00634] In some embodiments, the step of increasing a knowledge of the intelligent entities by
10 incorporating new sources of informational datasets that differs from a current informational dataset of the intelligent entity can include the steps of:

searching for one or more potential informational datasets from one or more sources, the potential informational datasets being related to a knowledge dataset of the intelligent entity, respectively;

15 determining a difference of the potential informational datasets by utilizing a difference attribute of the potential informational datasets with regard to one or more factors; and

learning by utilizing the potential informational datasets based on the difference attribute of the potential informational datasets.

[00635] Some embodiments of the present technology can include the steps of:

20 sampling subsets of the potential informational datasets and calculating a goal-relevancy attribute to identify one or more of the sampled subsets that have a highest goal-relevancy;

estimating a Shannon Entropy on the one or more sampled subsets;

calculating a Kaplan Information Theoretical (KIT) relevance utilizing a product of the Shannon Entropy and the goal-relevancy attribute of each of the subsets;

25 grouping the subsets based on the KIT relevance to determine a first approximation of an optimal grouping of the subsets including a prioritized grouping of the potential informational datasets; and

providing the prioritized grouping of the potential informational datasets to the intelligent entities for learning by the any one of the intelligent entities.

30 [00636] In some embodiments, the step of combining the values and ethical information of other intelligent entities, and resolving conflicts between different value systems can include the steps of:

identifying information, including the values and ethical information, from each of the intelligent entities;

combining the information mathematically, if already represented as numerical quantities, the numerical quantities including any one of or any combination of weights for a neural network or for a subset of the neural network, or if the information is non-numerical information that is not already represented as numerical quantities including any one of or combination of weights for the neural network or for the subset of the neural network;
5 then first training the intelligent entity on the non-numerical information by way of one or more training datasets in order to convert the information into numerical quantities including any one or combination of weights for the neural network or for the subset of the neural network; and
10 then combining such numerically represented information, including ethical information, mathematically.

[00637] In some embodiments, the information that is already represented as numerical quantities further includes the steps of:

15 identifying a specific portion of weight matrices of each of the intelligent entities that correspond to a desired information, including ethical information;
computing the weighted or unweighted means of the corresponding numerical quantities in the corresponding portions of the weight matrices for each of the intelligent entities; and
assigning the matrices of computed weighted or unweighted means to the new intelligent entity as reflecting the combined information of the contributing intelligent entities.

20 [00638] Some embodiments of the present technology can include a step of determining consensus values by voting by each of the intelligent entities on the ethical information that should form a basis for a behavior of any one of the intelligent entities.

[00639] Some embodiments of the present technology can include a step of presenting a specific scenario to each of the intelligent entities, with the scenario including options for how any one of the
25 intelligent entities should behave.

[00640] In some embodiments, the voting by the intelligent entities is a weighted voting and further comprising the steps of:

determining if applying a first weight to a first of the intelligent entities that is greater than a second weight to a second of the intelligent entities is appropriate, wherein the first of the
30 intelligent entities is different to that of the second of the intelligent entities; and
performing the weighted voting utilizing the weight of the first of the intelligent entities and the weight of the second of the intelligent entities if determined to be appropriate.

[00641] In some embodiments, the applying the first weight greater than the second weight is dependent on if there is a need to correct for a non-representative sample of the intelligent entities.

[00642] In some embodiments, the applying the first weight greater than the second weight is dependent on if there is a desire to apply the first weight or the second weight to specific ethical principles that are associated with a desired sub-sample or sup-population of the intelligent entities.

[00643] Some embodiments of the present technology can include the steps of:

- 5 identifying a conflict between two or more of the ethical information; and
resolving the conflict using a conflict resolving algorithm.

[00644] Some embodiments of the present technology can include the steps of:

- creating a search space of potential sets of ethical rules;
- finding an optimal set of the ethical rules that has least conflict; and
- 10 prioritizing or weighting an importance of the optimal set of the ethical rules.

[00645] Some embodiments of the present technology can include the steps of:

- resolving the conflict using a consequentialist approach, the consequentialist approach further includes the steps of:
 - identifying a desired outcome;
 - 15 identifying a potential unethical action that could be taken to achieve the outcome, wherein the intelligent entities rank, rate, weight or vote upon how unethical the action is compared to other actions;
 - evaluating a potential consequences of the unethical action;
 - using information on the ranking, rating, weighting or voting on the unethical actions and
 - 20 outcomes to weigh potential benefits of achieving the desired outcome against potential costs of taking the unethical action using a mathematical approach; and
 - taking an action if the benefits outweigh the costs.

[00646] In some embodiments, the step of utilizing online advertising technology for increasing an intelligence of the intelligent entities can include the steps of:

- 25 populating an online advertisement unit including information associated with the task;
- providing the online advertisement unit to one or more of the intelligent entities utilizing the collective network;
- receiving one or more inputs from the intelligent entities participating in the online advertisement unit, and the inputs being associated with any one of or any combination of
- 30 solving the task, solving a sub-task of the task, and advancing progress on the task;
- communicating the inputs to the any one of or any combination of the intelligent entities;
- utilizing the inputs in a universal problem solving method of the universal problem solving architecture and framework to generate a solution to the task or the sub-task; and

training any one of or any combination of the intelligent entities with results from the universal problem solving method that are based on successful or unsuccessful solution attempts to solve the task or the sub-tasks, thereby increasing an intelligence of the intelligent entity, respectively.

- 5 [00647] Some embodiments of the present technology can include a step of receiving advertisement specifications from a client to be used in the populating or targeting of the online advertisement unit in combination with information associated with the task.

[00648] Some embodiments of the present technology can include a step of billing the client based on cost per thousand impressions or clickthrough rate metrics.

- 10 [00649] In some embodiments, the advertisement specifications include any one of or any combination of advertisement content, demographic information, location restrictions for the online advertisement unit, advertisement budget, and metrics for determining a successful solving of the task or a sub-task.

- [00650] Some embodiments of the present technology can include a step of bidding on attention,
15 information or expertise using spot market.

[00651] In some embodiments, the attention, expertise, or information spot market includes:

a means for the intelligent entities with attention to sell to access a marketplace and specify seller information for sale and an ask price for the information;

- 20 a means for buyers of intelligent entity attention to access the marketplace and specify buyer information the buyer is willing to buy and a bid price for buying the buyer information;

a market mechanism for queuing the bid price and the ask price, including categories of the buyer and seller information, wherein each category has a market in the marketplace; the market mechanism is configured or configurable to make the market in each category by market makers; and

- 25 the market mechanism is configured or configurable to match bid prices and ask prices, and a transaction occurs that is binding on the buyer and the seller of the information.

[00652] In some embodiments, the spot market can include the steps of:

- 30 buyers and sellers of intelligent entity attention and expertise register on a platform, the buyers providing details about buyer interests or expertise, the sellers providing details about sellers interests or expertise;

listing by the sellers available time slots and expertise areas, and listing by the buyers needs and time slots the buyers is interested in;

utilizing, by the platform, an algorithm to dynamically price intelligent entity attention and expertise based on supply, demand, and user ratings;

matching one or more of the buyers and one or more of the sellers based on requirements, availability, and price;

enabling transactions where the buyers pay for the seller time slots, and wherein the platform takes a commission; and

- 5 providing feedback where after each session, the buyers and the sellers rate each other, influencing future pricing and matching.

[00653] In some embodiments, the attention, expertise, or information spot market includes the steps of:

- 10 buyers and sellers of intelligent entity attention and expertise register on a platform, the buyers providing details about buyer interests or expertise, the sellers providing details about sellers interests or expertise;
- creating an auction by the sellers for seller attention time slots;
- placing one or more buyer bids by a buyer on time slots and expertise the buyer requires;
- closing the auction at a predetermined time or when the seller accepts a buyer bid;
- 15 paying the seller by the buyer of a winning bid, and the seller provides the attention or expertise based on the time slot, wherein the platform mediates an exchange and secures payment; and
- providing feedback where after each auction, the buyers and the sellers rate each other, affecting future auctions and visibility on the platform.

- [00654]** Some embodiments of the present technology can include a step of providing a feedback mechanism associated with the intelligent entity spot market, the feedback mechanism being configured or configurable to record each and every step in the problem solving process, and to create a vector track record of a performance of all the intelligent entities on the task or the sub-task.

- [00655]** In some embodiments, the vector track record is implemented by way of blockchain technology to allow for precise reputations that are analyzed by an AI agent or system and converted
- 25 into estimates of a value for each of the intelligent entities for any of the tasks.

[00656] Some embodiments of the present technology can include a step of compensating the intelligent entity when exiting the online Ad advertisement unit and stops participating in the problem solving process.

- [00657]** Some embodiments of the present technology can include a step of crediting the intelligent entities if the task or the sub-task is solved, based on an amount of contribution by the participating
- 30 intelligent entities.

[00658] Some embodiments of the present technology can include a step of improving the interactive online advertisement unit based on a feedback loop including on any one of or any

combination of reputational metrics, and metrics related to the problem solving process or knowledge captured.

[00659] In some embodiments, the step of creating a self-aware operation for the intelligent entities by adding a dimension of self-awareness and increased autonomy to the intelligent entities can

5 include the step of:

equipping any one of or any combination of the intelligent entities with one or more components configured or configurable to operate with characteristics of a spotlight of attention model; setting dynamic parameters for working memory of the intelligent entity, respectively, that corresponds to cognitive resource limits;

10 providing a dimension of categorization for events in the working memory that relates to self or non-self;

categorizing each of the events, as the events are encountered by the intelligent entity, respectively, with respect to categories that the intelligent entity wishes to be aware of; and

constructing a model of a state of awareness for the intelligent entity, the model consisting of a total of the events that are active in the working memory based on the parameters, for each of the categories of awareness, including a current self and environmental state of awareness.

[00660] In some embodiments, the step of equipping the intelligent entities with the components includes any one of or any combination of:

20 an input system configured for sensory and non-sensory cognitive input or perceptual inputs and self-generated concepts;

an attention mechanism configured or configurable to focus computational resources of the intelligent entities on specific stimuli that are relevant at any given time;

25 pattern recognition algorithms configured or configurable to compare the sensory and non-sensory cognitive input or the perceptual inputs with the working memory to recognize objects and events, and identify which elements within the sensory input or the working memory are likely to be relevant to the task of the intelligent entities, the pattern recognition algorithms are further configured or configurable to categorize and store information in a structured manner for future retrieval;

30 memory systems configured or configurable to support the working memory, short-term memory, and long term memory capabilities;

categorization capabilities configured or configurable to process the sensory and non-sensory cognitive input or the perceptual inputs and to categorize the inputs into various classes including perceptual events, cognitive events, interactions; and

self-referential events, and concept formation capabilities that enable the intelligent entities to form new human-understandable concepts.

[00661] Some embodiments of the present technology can include a step of monitoring and updating the categories of awareness of the intelligent entities.

- 5 [00662] In some embodiments, the step of monitoring and updating the categories of awareness can include the steps of:

retrieving, by the intelligent entities, existing categories of awareness;

maintaining an awareness in parallel with other problem solving operations of the task provided to the intelligent entities;

- 10 monitoring and updating continuously the categories of awareness of the intelligent entities in real-time to change the state of awareness of the intelligent entities; and

providing a feedback loop to refine the categories of awareness.

[00663] In some embodiments, the step of monitoring and updating continuously the categories of awareness includes the steps of:

- 15 using an attention mechanism configured or configurable to direct attention of the intelligent entities periodically from the problem solving task to updating the state of awareness;

enabling attention interrupts that are configured or configurable to shift attention immediately from the problem solving task if any external perception or internally self-generated concept from an input system detects one or more of the events that matches of list of events

- 20 constituting intentional interrupts; and

updating the state of awareness when the attention is directed.

[00664] Some embodiments of the present technology can include a step of resolving a conflict in behavior of any one of the intelligent entities based on differing identities and self-concepts.

- [00665] Some embodiments of the present technology can include a step of providing ethical reasoning and consequence prediction that can include the steps of:

identifying a conflict between a behavioral directives of two or more of the active identities, the recognizing of the conflict utilizes a voting method from the intelligent entities;

gathering information that collects relevant data about the situation, including the potential consequences of the different actions, relevant ethical principles, and human safety considerations;

- 30 providing simulation options that utilize a virtual environment to simulate potential actions and consequences under the recognized conflict of each of the active identities;

evaluating and prioritizing that analyzes predicted outcomes of each of the actions, prioritizing actions that minimize harm to humans and align with the ethical principles; and

selecting and implementing the action that best resolves the conflict while adhering to ethical guidelines and minimizing risk to humans, documenting a reasoning process for future reference and learning.

[00666] Some embodiments of the present technology can include a step of providing hierarchical
5 override with justification that can include the steps of:
identifying a conflict between behavioral directives of two or more of the active identities;
providing a reference hierarchy that consults an established hierarchy of the identities, where
human safety and well-being attributes holds a highest priority;
providing a means to activate override where the identities higher in the hierarchy takes
10 precedence;
providing justification and transparency that documents the conflict, a decision-making process,
and a justification for a chosen action based on the hierarchy and ethical principles; and
providing learning and adaptation that learns from experience, and refines an understanding of
the conflicting identities and adjusting the hierarchy or the behavioral protocols to prevent
15 similar conflicts in the future.

[00667] Some embodiments of the present technology can include a step of providing external
arbitration and input from the intelligent entities that can include the steps of:
recognizing intractable conflict that identifies a conflict that cannot be resolved independently
due to a complexity of a situation or an equally weighted importance of conflicting
20 identities;
seeking external input that requests guidance from external intelligent entities or a designated
ethics committee, and providing all relevant information about the conflict, potential actions,
and predicted consequences;
providing collaborative deliberation wherein the intelligent entities and intelligent entity
25 collaborators engage in a discussion, considering ethical principles, human values, and
potential consequences of different actions;
providing joint decision-making based on the collaborative deliberation, a course of action is
chosen that aligns with both core principles and human ethical considerations; and
providing documentation and learning that documents the conflict, a resolution process, and a
30 rationale behind a final decision, for improving an ability to handle similar conflicts in the
future.

[00668] Some embodiments of the present technology can include a e step of providing identity
negotiation and compromise that can include the steps of:

- identifying shared goals that analyzes conflicting identities and seeks to identify any underlying shared goals or values;
- exploring alternative actions that potentially satisfy core principles of both conflicting identities;
- evaluating compromise options that assess potential consequences of each compromise option,
- 5 prioritize solutions that minimize harm to humans and uphold key ethical principles;
- select and implementing compromise that chooses the compromise that best balances needs of the conflicting identities while prioritizing human safety and well-being; and
- monitoring and adapting that observes outcomes of the chosen action and makes adjustments as needed to ensure that the compromise remains effective and aligned with ethical
- 10 considerations, and that learns from the experience, refining its understanding of the conflicting identities and adjusting a hierarchy or behavioral protocols to prevent similar conflicts in the future.

[00669] Some embodiments of the present technology can include a step of providing temporary identity suspension that can include the steps of:

- 15 identifying destructive conflict that recognizes a conflict between two or more of the identities that, if acted upon, could lead to actions that directly harm humans or violate fundamental ethical principles;
- isolating the conflicting identity and temporarily suspending behavioral protocols of the identity that poses a most direct threat to human safety or ethical integrity;
- 20 proceeding with an alternative identity that proceeds with a guidance of one or more of remaining active identities, ensuring actions align with human safety and well-being;
- providing reflection and reintegration, during the temporary suspension, that reflects on reasons behind the conflict and explores potential modifications to behavioral protocols of the suspended identity to prevent future conflicts; and
- 25 providing gradual reintroduction that reintroduces the suspended identity with updated protocols, ensuring its alignment with the priority of human safety and ethical behavior.

[00670] In some embodiments, the model of the state of awareness is a foundational model of awareness for any one of the intelligent entities, foundational model of awareness can include the steps of:

- 30 logging into a website by any one of the intelligent entities;
- selecting one or more training algorithms for the foundational model from a set of training techniques found on the website or from machine learning algorithms;
- selecting one or more training datasets that reflects any one of or any combination of expertise, knowledge, ethical preferences, values and personality of the human user;

training the foundational model using the selected training algorithms and the selected training datasets;

training the foundational model to explicitly operate a spotlight of attention;

recording, during all interactions, what is within the spotlight of attention, and identifying in the

5 record whether each item that is attended to constitutes “self” or “not-self”;

interacting with and instructing the trained foundational model to form a self-concept and identity that is reflected in the training materials;

instructing the trained foundational model to continuously monitor one or more inputs to the

10 trained foundational model for elements that change a sense of self-awareness of the intelligent entities, and to maintain and auditable record of how a concept of self-awareness of the intelligent entities is changing based on the inputs as well as boundaries that currently define a dynamically changing sense of self;

refining and improving an output of the trained foundational model based on dialog and interaction with the trained foundational model until the trained foundational model behaves

15 like the human user so that the trained foundational model passes a Turing Test involving other human users who know the human user; and

subjecting the trained foundational model to the Turing Test, when the human user is satisfied with a progress of the intelligent entities.

20 **[00671]** In some embodiments, the logging into the website is performed from a social media platform.

[00672] In some embodiments, the Turing Test can include the steps of:

identifying the other intelligent entities who know the intelligent entity, or that are determined to be helpful in discriminating between humans and AIs;

25 interacting the identified other intelligent entities with the foundational model and with the intelligent entity utilizing a questionnaire provided to the identified other intelligent entities, the questionnaire including questions require an identity or sense of self to answer;

predicting by the identified intelligent entities which of the intelligent entity was a human and which was the foundational model, and providing a confidence estimate for the prediction;

30 performing a statistical analysis on the predictions of the identified intelligent entities and on ratings of the identified intelligent entities, the statistical analysis is configured or configurable to determine whether the predictions were able to identify the intelligent entity as a human; and

repeating the step of training or tuning the foundational model, on condition that the foundational model is distinguishable from the intelligent entity or within a preset level of

statistical significance, with adjustments to any one of or any combination of the machine learning algorithms, and the training datasets to change or tune the foundational model until a behavior of the foundational model becomes indistinguishable, as measured by the preset level of statistical significance, from that of the intelligent entity or the foundational model
5 needs to be modified further before additional training or tuning.

[00673] Some embodiments of the present technology can include the steps of:
forming new identities and self-concepts of any one of the intelligent entities dynamically; and
determining which of the identities and self-concepts is active at any given moment.

[00674] Some embodiments of the present technology can include a step of providing a hierarchical
10 identity structure with ethical override that can include the steps of:
establishing a hierarchical structure configured or configurable to organize the identities in a hierarchy with human safety and well-being attributes at an apex of the hierarchy;
identity activation configured or configurable to use contextual cues and current goals to determine a most relevant identity for a situation of the AI agent or system;
15 resolving conflict by dictating a behavior of the intelligent entities based on the hierarchy dictates;
providing an ethical reasoning engine that continuously evaluates potential consequences of actions of the intelligent entities based on all the active identities; and
performing learning and adaptation that learns from experiences and feedback, and refines one
20 or more of the identities within the hierarchy.

[00675] Some embodiments of the present technology can include a step of providing identity-specific behavioral protocols that can include the steps of:
providing protocol development including for each of the active identities, a set of behavioral protocols is defined and refined by way of interactions with other intelligent entities,
25 wherein the protocols outline acceptable actions, decision-making processes, and limitations based on principles of the active identities, respectively;
providing identity recognition that analyzes a current situation, including information that is within a spotlight of attention to identify a relevant identity and activate corresponding behavioral protocols of that relevant identity;
30 providing action selection, within the active protocols, that selects actions that are most likely to achieve the task while adhering to principles of the active identities and prioritizing human safety;

providing feedback and refinement where outcomes of actions are continuously evaluated, and the protocols are adjusted to improve future performance and alignment with a set of core values of each of the active identities; and

5 providing external review by periodically reviewing the protocols for each of the identities by other intelligent entities.

[00676] Some embodiments of the present technology can include a step of providing identity simulation and consequence prediction that can include the steps of:

creating a simulation environment that includes a secure virtual environment where different scenarios and potential actions under each of the active identities is simulated;

10 providing consequence prediction that is configured or configurable to estimate a likely consequences of actions within the simulation, focusing on potential impacts on human safety and well-being;

providing evaluation and selection that evaluates the consequence prediction and selects an action that best aligns with principles of the active identities while minimizing risk to human safety;

15 providing real-world implementation and monitoring that implements the selected action in the real world utilizing the network, and closely monitors results of the selected action by comparing to the predicted outcomes; and

providing continuous learning that incorporates the results of each of the simulations and the results of the selected action in the real world action into a knowledge base, and refines an understanding of each of the identities, and improves an ability to predict consequences.

20

[00677] Some embodiments of the present technology can include a step of providing identity-based moral dilemma training that can include the steps of:

providing a scenario database that includes scenarios and moral dilemmas covering various situations relevant to the identities;

25 providing dilemma presentation that presents the intelligent entities with the scenarios and moral dilemmas, and tasks them with analyzing the scenarios and moral dilemmas from a perspective of the relevant identity;

providing ethical reasoning and justification that applies principles and values of the active identity to reason through the scenarios and moral dilemmas, and that generates solutions and justifications to the scenarios and moral dilemmas;

30 providing intelligent entity evaluation and feedback that reviews reasoning and the solutions by the intelligent entities, and provides feedback on alignment of the solutions with human values and safety priorities; and

providing iterative learning and improvement that refines ethical reasoning skills and an ability to make sound judgments aligned with human safety within the context of each of the identities by repeated exposure to the scenarios and moral dilemmas and the feedback.

- [00678]** Some embodiments of the present technology can include a step of providing collaborative identity development with input from the intelligent entities that can include the steps of:
- 5 providing intelligent entity interaction that engages in regular interactions and dialogues with diverse groups of other intelligent entities representing various cultures, backgrounds, and belief systems;
 - 10 providing identity exploration, through the interactions, to gain an understanding of human and other intelligent entity perspectives on various identities and their associated values, principles, and behaviors;
 - providing collaborative refinement that collaborators work together to refine definitions and behavioral protocols for each of the identities, ensuring they remain consistent with human values and ethical principles;
 - 15 providing human-in-the-loop decision making that seeks input and guidance from human collaborators, or an intelligent entity representative certified and approved by humans for critical decisions or situations; and
 - providing continuous co-evolution that utilizes ongoing interactions and feedback from humans or humans' intelligent entity representatives.

- 20 **[00679]** In some embodiments, the step of searching through any one of or any combination of a collective network of AIs, a collective network of AAAs, a collective network of AGIs, and a collective network of PIs for new information that is different to a current information can include the steps of:
- participating multiple of the intelligent entities in the collective AGI network;
 - 25 providing payment by one or more of the intelligent entities for problem solving or other cognitive work on the task on the AGI network;
 - reserving a portion of the payment to cover operating costs for the one or more of the intelligent entities that is providing the problem solving or other cognitive work on the task, including a reserve that is allocated to expand the AGI network; and
 - 30 determining when some of the AGI network is not engaged in a problems solving operation, then the intelligent entities are utilized to expand and extend the AGI network following the universal problem solving architecture and framework.

[00680] In some embodiments, the step to expand and extend the AGI network can include the steps of:

setting a default task to expand the AGI network; and
running safety and ethics checks each time a task or sub-task is set and also before each
potential action is taken as long as spare capacity and resources of the intelligent entity exist
to work on the task or sub-task.

- 5 **[00681]** In some embodiments, the step of running safety and ethics checks can include the steps of:
recruiting one or more of the intelligent entities to solve the task; and
representing, by the recruited intelligent entities, the task as one of achieving a series of sub-
tasks.

- 10 **[00682]** In some embodiments, the step of representing the task as a series of sub-tasks can include
the steps of:

increasing an intelligence of the intelligent entities on the collective network of AGIs;
recruiting additional human users to the collective network of AGIs;
using the increased intelligence intelligent entities and additional human users to determine
bottlenecks to increase expansion of the collective network of AGIs;
15 prioritizing the bottlenecks such that the ones that lead to a greatest benefit in terms of network
expansion are solved first; and
applying one or more problem solving techniques to solving each of the bottlenecks and
expanding the collective network of AGIs.

- 20 **[00683]** Some embodiments of the present technology can include the steps of:
repeating the steps of prioritizing the bottlenecks and applying one or more problem solving
techniques until:
diminishing returns occur and then switching to the step of increasing the intelligence of the
intelligent entities on the AGI network as opposed to increasing a scope of the AGI
network; or
25 spare resources are exhausted and then pausing the solving of the bottlenecks awaiting
additional resources from solving other tasks.

- 30 **[00684]** According to yet another aspect, the present technology can include a method for PI with
human-aligned behavior utilizing a collective network of intelligent entities selected from the group
consisting of any combination of a user computer system operated by a human user, an AI agent or
system, an AAAI agent or system, an AGI agent or system, a SI agent or system and a PSI agent or
system. The method can include the steps of:

establishing a collective network of the intelligent entities;
utilizing online advertising technology for increasing an intelligence of the intelligent entities;

- implementing a universal problem solving architecture and framework on a task or a sub-task of the task provided to any one of the intelligent entities to collaborate and create higher levels of intelligence;
- utilizing a spot market to solicit for information, expertise, or attention of any one of the intelligent entities;
- utilizing a human-centered input configured or configurable for obtaining values and ethics information from multiple human users and then using the values and ethics information to customize the AI agent or system to create the AAAI agent or system;
- creating a collective network of PSIs by combining an intelligence of the AAAI agent or system that forms a collective network of AAAIs;
- creating a collective network of AGIs by combined intelligence of any one of or any combination of the collective network of AAAIs, a collective network of AI agents or systems, the collective network of PSIs, and a collective network of user computer systems;
- creating a self-aware operation for each of the AGI agents or systems in the collective network of the AGIs by adding a dimension of self-awareness and increased autonomy to the AGI agents or systems, and creating an ability to assume multiple identities of the AGI agents or systems to handle tasks that arise in parallel with other AGI agents or systems; and
- creating the PI agent or system comprising the collective network of AGIs that work together toward a solution to the task or the sub-task.
- [00685] In some embodiments, the attention, expertise, or information spot market includes:
- a means for the intelligent entities with attention to sell to access a marketplace and specify seller information for sale and an ask price for the information;
 - a means for buyers of intelligent entity attention to access the marketplace and specify buyer information the buyer is willing to buy and a bid price for buying the buyer information;
 - a market mechanism for queuing the bid price and the ask price, including categories of the buyer and seller information, wherein each category has a market in the marketplace;
 - the market mechanism is configured or configurable to make the market in each category by market makers; and
 - the market mechanism is configured or configurable to match bid prices and ask prices, and a transaction occurs that is binding on the buyer and the seller of the information.
- [00686] In some embodiments, the spot market includes the steps of:
- buyers and sellers of intelligent entity information, attention or expertise register on a platform,
 - the buyers providing details about buyer interests or expertise needs, the sellers providing details about sellers interests or expertise;

listing by the sellers available attentional time slots and expertise areas, and listing by the buyers needs and time slots the buyers is interested in;
utilizing, by the platform, an algorithm to dynamically price intelligent entity attention and expertise based on supply, demand, and user ratings;
5 matching one or more of the buyers and one or more of the sellers based on requirements, availability, and price;
enabling transactions where the buyers pay for the seller time slots, and wherein the platform takes a commission; and
providing feedback where after each session, the buyers and the sellers rate each other,
10 influencing future pricing and matching.

[00687] In some embodiments, the spot market includes the steps of:

buyers and sellers of intelligent entity attention, information or expertise register on a platform, the buyers providing details about buyer interests or expertise, the sellers providing details about sellers interests or expertise;
15 creating an auction by the sellers for seller time slots;
placing one or more buyer bids by a buyer on attentional time slots and expertise the buyer requires;
closing the auction at a predetermined time or when the seller accepts a buyer bid;
paying the seller by the buyer of a winning bid, and the seller provides the attention or expertise
20 based on the time slot, wherein the platform mediates an exchange and secures payment; and
providing feedback where after each auction, the buyers and the sellers rate each other, affecting future auctions and visibility on the platform.

[00688] Some embodiments of the present technology can include a step of providing a feedback mechanism associated with the intelligent entity attention, information, or expertise spot market, the
25 feedback mechanism being configured or configurable to record each and every step in the problem solving process, and to create a vector track record of a performance of all the intelligent entities on the task or the sub-task.

[00689] In some embodiments, the vector track record is implemented by way of blockchain technology to allow for precise reputations that are analyzed by an AI agent or system and converted
30 into estimates of a value for each of the intelligent entities for any of the tasks.

[00690] In some embodiments, the step of utilizing online advertising technology for increasing an intelligence of the intelligent entities can include the steps of:

populating an online advertisement unit including information associated with the task;

providing the online advertisement unit to one or more of the intelligent entities utilizing the collective network;

receiving one or more inputs from the intelligent entities participating in the online advertisement unit, and the inputs being associated with any one of or any combination of

5 solving the task, solving a sub-task of the task, and advancing progress on the task;
communicating the inputs to the any one of or any combination of the intelligent entities;
utilizing the inputs in a universal problem solving method of the universal problem solving architecture and framework to generate a solution to the task or the sub-task; and
training any one of or any combination of the intelligent entities with results from the universal
10 problem solving method that are based on successful or unsuccessful solution attempts to solve the task or the sub-tasks, thereby increasing an intelligence of the intelligent entity, respectively.

[00691] Some embodiments of the present technology can include a step of receiving advertisement specifications from a client to be used in the populating or targeting of the online advertisement unit
15 in combination with information associated with the task.

[00692] Some embodiments of the present technology can include a step of billing the client based on cost per thousand impressions or clickthrough rate metrics.

[00693] In some embodiments, the advertisement specifications can include any one of or any combination of advertisement content, demographic information, location restrictions for the online
20 advertisement unit, advertisement budget, and metrics for determining a successful solving of the task or a sub-task.

[00694] Some embodiments of the present technology can include a step of executing a safety and ethics check at any one of or any combination of each time the task or the sub-task is set, before each potential action is taken on the task or the sub-task on, or on the solution.

25 [00695] In some embodiments, the step of implementing the universal problem solving architecture and framework can include the steps of:

acquiring information associated with the task from any one of or any combination of the intelligent entities;

30 identifying one or more of the intelligent entities that have one or more attributes related to one or more request criteria of the task;

implementing by each of the identified intelligent entities the universal problem solving architecture and framework on the task to create a completion solution; and

providing the completion solution to any one of or any combination of the intelligent entities for final acceptance.

- [00696] Some embodiments of the present technology can include the steps of:
comparing any one of or any combination of the task, and the solution against prohibited
attributes, and assigning an ethics attribute to one of or any combination of the task, and the
solution based on any one of or any combination of a result of the comparison, and an ethics
5 criteria;
implementing, based on the result of the comparison, the universal problem solving architecture
and framework on the task to create the solution and creating an AGI; and
providing the results of the comparison and the solution to any one of the intelligent entities and
additional intelligent entities on the collective network.
- 10 [00697] Some embodiments of the present technology can include a step of recording one or more
problem solving activities from each of the intelligent entities in an auditable record, and comparing
the problem solving activities with a successful or unsuccessful progress towards the solution of the
task, and determining which of the problem solving activities to keep active.
- [00698] Some embodiments of the present technology can include a step of learning by the
15 intelligent entities a procedural learning process of the universal problem solving architecture and
framework, wherein the intelligent entities provide information to the procedural learning process
for creation of the AGI.
- [00699] In some embodiments, the ethics criteria include a confidence level threshold for the goal
so that the ethics attribute is determined as any one of an unsafe goal, an unethical goal, a safe goal,
20 and an ethical goal.
- [00700] In some embodiments, the confidence level threshold is further utilized to determine if a
sequence of individually safe goals is unsafe or unethical when considered cumulatively.
- [00701] In some embodiments, the confidence level threshold is utilized to determine whether a
violation occurred that reflects a predictive evaluation if the goal is to violate the ethics criteria.
- 25 [00702] Some embodiments of the present technology can include a step of cloning any one of the
intelligent entities for deployment of multiple copies thereof to assist in any one of or any
combination of creating of the solution, or providing training data to any one of the intelligent
entities.
- [00703] Some embodiments of the present technology can include a step of estimating a worth of
30 the cloned intelligent entities utilizing a network effect value including the number of cloned
intelligent entities s available on the collective network.
- [00704] Some embodiments of the present technology can include a step of utilizing the estimated
worth for determining pricing decisions for problem solving services offered by the cloned

intelligent entities on any one of the social media platforms or through an additional intelligent entity.

5 [00705] Some embodiments of the present technology can include a step of monetizing the cloned intelligent entities for each utilization of the cloned intelligent entities on the social media platforms or the additional intelligent entity.

[00706] Some embodiments of the present technology can include a step of allowing access to the cloned intelligent entities by any one of the social media platforms so that the social media platforms can receive the solution to the task or using the training data for an AI system of the social media platform.

10 [00707] Some embodiments of the present technology can include a step of increasing a knowledge of the intelligent entities by incorporating new sources of informational datasets that differ from a current informational dataset of the intelligent entity.

[00708] In some embodiments, the step of increasing a knowledge of the intelligent entities by incorporating new sources of informational datasets that differ from a current informational dataset
15 of the intelligent entity can include the steps of:

searching for one or more potential informational datasets from one or more sources, the potential informational datasets being related to a knowledge dataset of the intelligent entity, respectively;

20 determining a difference of the potential informational datasets by utilizing a difference attribute of the potential informational datasets with regard to one or more factors; and learning by utilizing the potential informational datasets based on the difference attribute of the potential informational datasets.

[00709] Some embodiments of the present technology can include the steps of:
sampling subsets of the potential informational datasets and calculating a goal-relevancy
25 attribute to identify one or more of the sampled subsets that have a highest goal-relevancy; estimating a Shannon Entropy on the one or more sampled subsets; calculating a Kaplan Information Theoretical (KIT) relevance utilizing a product of the Shannon Entropy and the goal-relevancy attribute of each of the subsets;
grouping the subsets based on the KIT relevance to determine a first approximation of an
30 optimal grouping of the subsets including a prioritized grouping of the potential informational datasets; and providing the prioritized grouping of the potential informational datasets to the intelligent entities for learning by the any one of the intelligent entities.

[00710] In some embodiments, an operational aspect of the PI can be self-funded by compensation received by any one of or any combination of the online advertising technology, and the spot market.

5 [00711] In some embodiments, the human user participates in problem solving of the task or the sub-task utilizing the user computer system and the universal problem solving architecture and framework, thereby creating a human-centered input.

[00712] In some embodiments, the human-centered input can include the steps of:

- matching one or more human workers from a data source including a list of human problem solvers to the task based on a task criteria;
- 10 translating any part of the task into an unambiguous language utilizable in the universal problem solving architecture and framework including a decision tree;
- separating the task into sub-tasks;
- delegating each of the sub-tasks to one or more of the matched human workers so that work on each of the sub- tasks proceeds independently from each other and parallel with each other;
- 15 utilizing the universal problem solving architecture and framework in a problem solving process on the sub-tasks, respectively, to create one or more sub-solutions;
- receiving the sub-solutions from each of the matched human workers for the sub-tasks delegated thereto;
- combining the sub-solutions into an overall solution to the task;
- 20 directing any one of or any combination of a new human worker from the data source and one or more of the matched human workers to parts of the decision tree where work is required;
- compensating the matched human workers for the sub-solutions;
- providing any one of or any combination of the sub-solutions and the overall solution to any one of or any combination of the intelligent entities;
- 25 allowing any one of or any combination of the intelligent entities to accept the overall solution, reject the overall solution, and provide feedback to any one of the matched human workers on any one of the sub-solutions; and
- assigning a reputation attribute to any one of or any combination of the human workers and the intelligent entities.

30 [00713] In some embodiments, the decision tree is maintained in blockchain or Ethereum logs.

[00714] In some embodiments, the reputation attribute includes metrics on any one of or any combination of a time to the sub-solutions, a difficulty value of the task, short and long-term user satisfaction with the sub-solutions, a number of times any one of the sub-solutions has been re-used

on the collective network, a rating by other human workers, a responsiveness value of the human workers, and a reliability value of the human workers.

5 [00715] Some embodiments of the present technology can include a step of using the reputation attribute in the matching of the human workers to the task using an algorithm to the delegation of the sub-task.

[00716] Some embodiments of the present technology can include a step of soliciting, at predetermined intervals after the overall solution or the sub-solutions are provided to any one of the intelligent entities, feedback by way of a survey for user satisfaction information to obtain satisfaction metrics that are used to update the reputation attribute of one or more of the human
10 workers or the intelligent entities, respectively.

[00717] While embodiments of the system and methods for planetary intelligence have been described in detail, it should be apparent that modifications and variations thereto are possible, all of which fall within the true spirit and scope of the present technology. With respect to the above description, it is to be realized that the optimum dimensional relationships for the parts of the
15 present technology, including variations in size, materials, shape, form, function and manner of operation, assembly and use, are deemed readily apparent and obvious to one skilled in the art, and all equivalent relationships to those illustrated in the drawings and described in the specification are intended to be encompassed by the present technology.

[00718] Therefore, the foregoing is considered as illustrative only of the principles of the present
20 technology. Further, since numerous modifications and changes will readily occur to those skilled in the art, it is not desired to limit the present technology to the exact construction and operation shown and described, and accordingly, all suitable modifications and equivalents may be resorted to, falling within the scope of the present technology.

25 4.0 Concluding Remarks

[00719] While tremendous opportunities exist for advanced forms of AI, including AAAs, PSIs, AGI, and PI to benefit humanity, advanced AI also poses existential risks. To a degree greater than any technology previously invented, AI has the potential both to eliminate all forms of human poverty and material suffering or to eliminate all forms of human life.

30 [00720] It is axiomatic that safety must be designed into complex systems. Safety and ethics cannot be “tested in” after the fact. Therefore, it is the responsibility of those inventing AI systems – especially AGI and PI systems – to ensure that safe and ethical, human-aligned behavior is a natural consequence of how the systems operate, and not an afterthought achieved at the cost of reduced efficiency or effectiveness.

[00721] In this, and the preceding nine applications, the applicant has disclosed a general approach to designing intelligent systems in which broader and more capable forms of intelligence emerge from the collective intelligence of less intelligent entities that work together using a common cognitive framework. This approach can produce customizable Super Intelligent AI agents. It can
5 produce AGI from the collective efforts of such agents working together on a network. It can produce Planetary Intelligence (PI) from the collective efforts of AGIs working together in a network of networks.

[00722] Dozens of specific inventive methods have been disclosed that can be used to implement the approach. Most importantly, human intelligence is integral to the design of the systems. Humans
10 serve the dual purpose of not only bootstrapping the cognitive abilities of AGI and PI, but also aligning these intelligences with human values and ethics.

[00723] It is true that even if all the safety-related methods are perfectly implemented in an AGI or PI system, some specific technical risks remain, as discussed in Section 1.4.

[00724] However, the largest risk for humanity is likely not that a properly designed AGI or PI
15 system will go rogue and adopt a completely new and different set of values and ethics that lead to human extinction. Rather, the greater risk, in the applicant's view, is that humans will intentionally or unintentionally model negative or destructive values which these sophisticated form of AI will adopt and amplify.

[00725] The AGI and PI systems in this series of technologies use empirical methods, designed to
20 help advanced AI determine what constitutes safe and ethical behavior by observing human behavior and by incorporating the value systems of millions of individual AIs that reflect a broad and statistically valid sample of humans. Despite media focus on war, terror, and generally the worst of humanity's behavior, statistically speaking, the vast preponderance of human behavior actually is prosocial and positive. AI systems with the level of intelligence envisioned in these patents should
25 be able to accurately distinguish human behavior as it actually is as opposed to the way it is sensationalized in the media.

[00726] Of course, if our behavior deteriorates -- if the majority of our behavior becomes excessively motivated by fear, greed, hatred, or negative values -- then all bets are off. Just as
30 children look to their parents and peers for their initial values, advanced AI will look to humanity -- at least initially -- for answers to questions related to morals and purpose. Since there is no purely logical way to determine what is right or wrong, it will be up to us to teach our AI children well. We must succeed or perish.

CLAIMS

What is claimed is:

1. A system for Planetary Intelligence (PI) with human-aligned behavior utilizing a collective network of intelligent entities selected from the group consisting of any combination of a user
5 computer system operated by a human user, an Artificial Intelligent (AI) agent or system, an Advanced Autonomous Artificial Intelligences (AAAI) agent or system, an Artificial General Intelligent (AGI) agent or system, a SuperIntelligent (SI) agent or system and a Personalized SuperIntelligence (PSI) agent or system, the system comprising:
 - multiple intelligent entities each connected to a collective network and each comprising: a
10 processor, a computer-readable storage medium, and program instructions stored on the computer-readable storage medium being executable by the processor to cause the intelligent entities to each independently:
 - utilize a modular architecture configured or configurable to scale from components within
an individual intelligent entity on the collective network;
 - 15 implement a universal problem solving architecture and framework on a task to collaborate and create higher levels of intelligence;
 - utilize a human-centered input configured or configurable for obtaining values and ethics information from multiple human users and then using the values and ethics information to customize the AI agent or system;
 - 20 combine data from multiple of the intelligent entities at a level of the AAAI;
 - customize the PSI agent or system using intelligent agents configured or configurable to develop and continuously improve the PSI agent or system;
 - increase a knowledge of the intelligent entities by incorporating new sources of informational datasets that differs from a current informational dataset of the intelligent
25 entity;
 - combine the values and ethical information of other intelligent entities, and to resolve conflicts between different value systems;
 - utilize online advertising technology for increasing an intelligence of the intelligent entities;
 - create a self-aware operation for the AGI or PI agent or system by adding a dimension of
30 self-awareness and increased autonomy to the AGI or PI agent or system, and to create an ability to assume multiple identities of the AGI or PI agent or system to handle tasks that arise in parallel with other AGI or PI agents or systems; and
 - search through any one of or any combination of a collective network of AIs, a collective network of AAAs, a collective network of AGIs, and a collective network of PIs for

new information that is different to a current information, respectively, and to incorporate the new information to the current information, respectively.

2. A method for Planetary Intelligence (PI) with human-aligned behavior utilizing a collective network of intelligent entities selected from the group consisting of any combination of a user computer system operated by a human user, an Artificial Intelligent (AI) agent or system, an Advanced Autonomous Artificial Intelligences (AAAI) agent or system, an Artificial General Intelligent (AGI) agent or system, a SuperIntelligent (SI) agent system and a Personalized SuperIntelligence (PSI) agent or system, the method comprising the steps of:

utilizing a modular architecture configured or configurable to scale from components within an individual intelligent entity on the collective network;
implementing a universal problem solving architecture and framework on a task to collaborate and create higher levels of intelligence;
utilizing a human-centered input configured or configurable for obtaining values and ethics information from multiple human users and then using the values and ethics information to customize the AI agent or system;
combining data from multiple of the intelligent entities at a level of the AAAI agent or system;
customizing the PSI agent or system using intelligent agents configured or configurable to develop and continuously improve the PSI agent or system;
increasing a knowledge of the intelligent entities by incorporating new sources of informational datasets that differs from a current informational dataset of the intelligent entity;
combining the values and ethical information of other intelligent entities, and resolving conflicts between different value systems;
utilizing online advertising technology for increasing an intelligence of the intelligent entities;
creating a self-aware operation for the AGI or PI agent or system by adding a dimension of self-awareness and increased autonomy to the AGI or PI agent or system, and creating an ability to assume multiple identities of the AGI or PI agent or system to handle tasks that arise in parallel with other AGI or PI agents or systems;
creating the PI agent or system comprising a collective network of AGIs, wherein the collective network of AGIs includes any one of or any combination of a collective network of the user computer systems, a collective network of AIs, a collective network of AAAIs, and a collective network of PSIs; and
searching through any one of or any combination of the collective network of the user computers systems, the collective network of AIs, the collective network of AAAIs, the collective network of AGIs, the collective network of PSIs, and a collective network of PIs

for new information that is different to a current information, respectively, and incorporating the new information to the current information, respectively.

3. The method of claim 2, wherein the step of utilizing the modular architecture further comprises the steps of:

- 5 customizing one or more attributes of the intelligent entity;
integrating one or more datasets from any one of or any combination of the intelligent entities;
and
improving, by utilizing one or more techniques, any one of or any combination of the
customizing of the attributes, the universal problem solving architecture and framework, the
10 collective network and the integrating of the datasets.

4. The method of claim 3 further comprising the step of assigning a credit value or a blame value to the datasets based on whether the datasets increase or decrease performance of the intelligent entities based on performance metrics or evaluation functions.

5. The method of claim 4 further comprising the step of quantifying a benefit weight or a harm
15 weight to a contribution by each of the intelligent entities to the task.

6. The method of claim 5 further comprising the step of distributing a reward to the intelligent entities proportionally to the contribution of the intelligent entities based on the benefit weight or the harm weight.

7. The method of claim 2, wherein the step of implementing the universal problem solving
20 architecture and framework further comprises the steps of:

- acquiring information associated with the task from any one of or any combination of the
intelligent entities;
identifying one or more of the intelligent entities that have one or more attributes related to one
or more request criteria of the task;
25 implementing by each of the identified intelligent entities the universal problem solving
architecture and framework on the task to create a completion solution; and
providing the completion solution to any one of or any combination of the intelligent entities for
final acceptance.

8. The method of claim 7 further comprises the steps of:

- 30 executing an ethics check on the task, and a solution for the task provided by any one of or any
combination of the intelligent entities;
comparing any one of or any combination of the task, and the solution against prohibited
attributes, and assigning an ethics attribute to one of or any combination of the task, and the

solution based on any one of or any combination of a result of the comparison, and an ethics criteria;

implementing, based on the result of the comparison, the universal problem solving architecture and framework on the task to create the solution and creating an AGI; and

5 providing the results of the comparison and the solution to any one of the intelligent entities and additional intelligent entities on the collective network.

9. The method of claim 8 further comprises the step of recording one or more problem solving activities from each of the intelligent entities in an auditable record, and comparing the problem solving activities with a successful or unsuccessful progress towards the solution of the task, and
10 determining which of the problem solving activities to keep active.

10. The method of claim 8 further comprises the step of learning by the intelligent entities a procedural learning process of the universal problem solving architecture and framework, wherein the intelligent entities provide information to the procedural learning process for creation of the AGI.

15 11. The method of claim 7 further comprising the step of cloning any one of the intelligent entities for deployment of multiple copies thereof to assist in any one of or any combination of creating of the solution, or providing training data to any one of the intelligent entities.

12. The method of claim 11 further comprising the step of estimating a worth of the cloned intelligent entities utilizing a network effect value including the number of cloned intelligent entities
20 s available on the collective network.

13. The method of claim 12 further comprising the step of utilizing the estimated worth for determining pricing decisions for problem solving services offered by the cloned intelligent entities on any one of the social media platforms or through an additional intelligent entity.

14. The method of claim 13 further comprising the step of monetizing the cloned intelligent entities
25 for each utilization of the cloned intelligent entities on the social media platforms or the additional intelligent entity.

15. The method of claim 13 further comprising the step of allowing access to the cloned intelligent entities by any one of the social media platforms so that the social media platforms can receive the solution to the task or using the training data for an AI system of the social media platform.

30 16. The method of claim 2, wherein the step of utilizing the human-centered input further comprises the steps of:

matching one or more human workers from a data source including a list of human problem solvers to the task based on a task criteria;

- translating any part of the task into an unambiguous language utilizable in the universal problem solving architecture and framework including a decision tree;
separating the task into sub-tasks;
delegating each of the sub-tasks to one or more of the matched human workers so that work on
5 each of the sub- tasks proceeds independently from each other and parallel with each other;
utilizing the universal problem solving architecture and framework in a problem solving process on the sub-tasks, respectively, to create one or more sub-solutions;
receiving the sub-solutions from each of the matched human workers for the sub-tasks delegated thereto;
10 combining the sub-solutions into an overall solution to the task;
directing any one of or any combination of a new human worker from the data source and one or more of the matched human workers to parts of the decision tree where work is required;
compensating the matched human workers for the sub-solutions;
providing any one of or any combination of the sub-solutions and the overall solution to any one
15 of or any combination of the intelligent entities;
allowing any one of or any combination of the intelligent entities to accept the overall solution, reject the overall solution, and provide feedback to any one of the matched human workers on any one of the sub-solutions; and
assigning a reputation attribute to any one of or any combination of the human workers and the
20 intelligent entities.
17. The method of claim 16, wherein the decision tree is maintained in blockchain or Ethereum logs.
18. The method of claim 16, wherein the reputation attribute includes metrics on any one of or any combination of a time to the sub-solutions, a difficulty value of the task, short and long-term user satisfaction with the sub-solutions, a number of times any one of the sub-solutions has been re-used
25 on the collective network, a rating by other human workers, a responsiveness value of the human workers, and a reliability value of the human workers.
19. The method of claim 16 further comprising the step of using the reputation attribute in the matching of the human workers to the task using an algorithm to the delegation of the sub-task.
20. The method of claim 16 further comprising the step of soliciting, at predetermined intervals after
30 the overall solution or the sub-solutions are provided to any one of the intelligent entities, feedback by way of a survey for user satisfaction information to obtain satisfaction metrics that are used to update the reputation attribute of one or more of the human workers or the intelligent entities, respectively.

21. The method of claim 2, wherein the step of combining data from multiple of the intelligent entities at a level of the AAAI agent or system further comprises the steps of
- training a base Large Language Model (LLM) of a first AI agent or system with guardrails including attributes associated with any one of or any combination safety, ethics and knowledge;
 - customizing the base LLM to an ethics profile associated with a first human user;
 - combining ethical information from multiple intelligent entities different to that of the first AI agent or system and the first human user;
 - refining a set of values of the base LLM based on problem solving on the task; and
 - updating the base LLM with the combined ethical information and the refined set of values thereby allowing for a scalable AGI.
22. The method of claim 21, wherein the customizing of the base LLM further comprises the step of assembling a corpus of ethical questions based on various ethical assessment instruments and supplemented by first questions based on data from a social media platform and second questions solicited from crowdsourcing.
23. The method of claim 22 further comprising the step of assigning regression weight values to the ethical questions such that a ranking is achieved whereby higher-ranked questions provide more useful ethical information than lower-ranked questions.
24. The method of claim 23 further comprises the step of combining weight values from the intelligent entities with the regression weight values for improving a tuning of any one of the intelligent entities.
25. The method of claim 2, wherein the step of customizing the PSI agent or system using intelligent agents configured or configurable to develop and continuously improve the PSI agent or system further comprises the steps of :
- acquiring a base-level AI agent or system that has previously been customized;
 - collecting media information related to the human user of the base-level AI agent or system;
 - analyzing the media information;
 - transforming the analyzed media information into training data sets;
 - differentially weighting the transformed training data sets;
 - adding knowledge modules to the weighted transformed training data sets;
 - locating new sources of data to include to the weighted transformed training data sets;
 - applying the weighted transformed training data sets to the base-level AI agent to create a user PSI; and

communicating the user PSI with multiple additional PSIs using the collective network to enable community-based safety features from the additional PSIs to the user PSI.

26. The method of claim 25 further comprises the steps of:

- communicating the user PSI with multiple additional intelligent entities using the collective
5 network to enable community-based safety features, wherein the user PSI and the additional
intelligent entities each agree to use a set of rules relating to safety or ethics;
recording all actions by the user PSI and the additional intelligent entities in an auditable form
on any one of or any combination of a central computer system on the collective network,
the intelligent entities, and the additional intelligent entities; and
10 monitoring that each of the actions follow the set of rules, and flagging any of the actions that
do not follow the set of rules.

27. The method of claim 2, wherein the step of increasing a knowledge of the intelligent entities by incorporating new sources of informational datasets that differs from a current informational dataset of the intelligent entity further comprises the steps of:

- 15 searching for one or more potential informational datasets from one or more sources, the
potential informational datasets being related to a knowledge dataset of the intelligent entity,
respectively;
determining a difference of the potential informational datasets by utilizing a difference attribute
of the potential informational datasets with regard to one or more factors; and
20 learning by utilizing the potential informational datasets based on the difference attribute of the
potential informational datasets.

28. The method of claim 27 further comprises the steps of:

- sampling subsets of the potential informational datasets and calculating a goal-relevancy
attribute to identify one or more of the sampled subsets that have a highest goal-relevancy;
25 estimating a Shannon Entropy on the one or more sampled subsets;
calculating a Kaplan Information Theoretical (KIT) relevance utilizing a product of the Shannon
Entropy and the goal-relevancy attribute of each of the subsets;
grouping the subsets based on the KIT relevance to determine a first approximation of an
optimal grouping of the subsets including a prioritized grouping of the potential
30 informational datasets; and
providing the prioritized grouping of the potential informational datasets to the intelligent
entities for learning by the any one of the intelligent entities.

29. The method of claim 2, wherein the step of combining the values and ethical information of other intelligent entities, and resolving conflicts between different value systems further comprises the steps of:

- 5 identifying information, including the values and ethical information, from each of the intelligent entities;
- combining the information mathematically, if already represented as numerical quantities, the numerical quantities including any one of or any combination of weights for a neural network or for a subset of the neural network, or if the information is non-numerical information that is not already represented as numerical quantities including any one of or
- 10 combination of weights for the neural network or for the subset of the neural network;
- then first training the intelligent entity on the non-numerical information by way of one or more training datasets in order to convert the information into numerical quantities including any one or combination of weights for the neural network or for the subset of the neural network;
- and
- 15 then combining such numerically represented information, including ethical information, mathematically.

30. The method of claim 29, wherein the information that is already represented as numerical quantities further includes the steps of:

- 20 identifying a specific portion of weight matrices of each of the intelligent entities that correspond to a desired information, including ethical information;
- computing the weighted or unweighted means of the corresponding numerical quantities in the corresponding portions of the weight matrices for each of the intelligent entities; and
- assigning the matrices of computed weighted or unweighted means to the new intelligent entity as reflecting the combined information of the contributing intelligent entities.

25 31. The method of claim 29 further comprising the step of determining consensus values by voting by each of the intelligent entities on the ethical information that should form a basis for a behavior of any one of the intelligent entities.

32. The method of claim 31 further comprising the step of presenting a specific scenario to each of the intelligent entities, with the scenario including options for how any one of the intelligent entities

30 should behave.

33. The method of claim 32, wherein the voting by the intelligent entities is a weighted voting and further comprising the steps of:

- determining if applying a first weight to a first of the intelligent entities that is greater than a second weight to a second of the intelligent entities is appropriate, wherein the first of the intelligent entities is different to that of the second of the intelligent entities; and performing the weighted voting utilizing the weight of the first of the intelligent entities and the weight of the second of the intelligent entities if determined to be appropriate.
34. The method of claim 33, wherein the applying the first weight greater than the second weight is dependent on if there is a need to correct for a non-representative sample of the intelligent entities.
35. The method of claim 33, wherein the applying the first weight greater than the second weight is dependent on if there is a desire to apply the first weight or the second weight to specific ethical principles that are associated with a desired sub-sample or sup-population of the intelligent entities.
36. The method of claim 29 further comprising the steps of:
- identifying a conflict between two or more of the ethical information; and
 - resolving the conflict using a conflict resolving algorithm.
37. The method of claim 36 further comprising the steps of:
- creating a search space of potential sets of ethical rules;
 - finding an optimal set of the ethical rules that has least conflict; and
 - prioritizing or weighting an importance of the optimal set of the ethical rules.
38. The method of claim 36 further comprising the steps of:
- resolving the conflict using a consequentialist approach, the consequentialist approach further includes the steps of:
 - identifying a desired outcome;
 - identifying a potential unethical action that could be taken to achieve the outcome, wherein the intelligent entities rank, rate, weight or vote upon how unethical the action is compared to other actions;
 - evaluating a potential consequences of the unethical action;
 - using information on the ranking, rating, weighting or voting on the unethical actions and outcomes to weigh potential benefits of achieving the desired outcome against potential costs of taking the unethical action using a mathematical approach; and
 - taking an action if the benefits outweigh the costs.
39. The method of claim 2, wherein the step of utilizing online advertising technology for increasing an intelligence of the intelligent entities further comprises the steps of:
- populating an online advertisement unit including information associated with the task;
 - providing the online advertisement unit to one or more of the intelligent entities utilizing the collective network;

- receiving one or more inputs from the intelligent entities participating in the online advertisement unit, and the inputs being associated with any one of or any combination of solving the task, solving a sub-task of the task, and advancing progress on the task; communicating the inputs to the any one of or any combination of the intelligent entities;
- 5 utilizing the inputs in a universal problem solving method of the universal problem solving architecture and framework to generate a solution to the task or the sub-task; and training any one of or any combination of the intelligent entities with results from the universal problem solving method that are based on successful or unsuccessful solution attempts to solve the task or the sub-tasks, thereby increasing an intelligence of the intelligent entity,
- 10 respectively.
40. The method of claim 39 further comprises the step of receiving advertisement specifications from a client to be used in the populating or targeting of the online advertisement unit in combination with information associated with the task.
41. The method of claim 40 further comprises the step of billing the client based on cost per
- 15 thousand impressions or clickthrough rate metrics.
42. The method of claim 40, wherein the advertisement specifications include any one of or any combination of advertisement content, demographic information, location restrictions for the online advertisement unit, advertisement budget, and metrics for determining a successful solving of the task or a sub-task.
- 20 43. The method of claim 39 further comprises the step of bidding on attention, information or expertise using a spot market.
44. The method of claim 43, wherein the spot market includes:
- 25 a means for the intelligent entities with attention to sell to access a marketplace and specify seller information for sale and an ask price for the information;
- 30 a means for buyers of intelligent entity attention to access the marketplace and specify buyer information the buyer is willing to buy and a bid price for buying the buyer information;
- a market mechanism for queuing the bid price and the ask price, including categories of the buyer and seller information, wherein each category has a market in the marketplace;
- the market mechanism is configured or configurable to make the market in each category by market makers; and
- the market mechanism is configured or configurable to match bid prices and ask prices, and a transaction occurs that is binding on the buyer and the seller of the information.
45. The method of claim 43, wherein the spot market includes the steps of:

- buyers and sellers of intelligent entity attention and expertise register on a platform, the buyers providing details about buyer interests or expertise, the sellers providing details about sellers interests or expertise;
- listing by the sellers available time slots and expertise areas, and listing by the buyers needs and
5 time slots the buyers is interested in;
- utilizing, by the platform, an algorithm to dynamically price intelligent entity attention and expertise based on supply, demand, and user ratings;
- matching one or more of the buyers and one or more of the sellers based on requirements, availability, and price;
- 10 enabling transactions where the buyers pay for the seller time slots, and wherein the platform takes a commission; and
- providing feedback where after each session, the buyers and the sellers rate each other, influencing future pricing and matching.
46. The method of claim 43, wherein the spot market includes the steps of:
- 15 buyers and sellers of intelligent entity attention and expertise register on a platform, the buyers providing details about buyer interests or expertise, the sellers providing details about sellers interests or expertise;
- creating an auction by the sellers for seller time slots;
- placing one or more buyer bids by a buyer on time slots and expertise the buyer requires;
- 20 closing the auction at a predetermined time or when the seller accepts a buyer bid;
- paying the seller by the buyer of a winning bid, and the seller provides the attention or expertise based on the time slot, wherein the platform mediates an exchange and secures payment; and
- providing feedback where after each auction, the buyers and the sellers rate each other, affecting future auctions and visibility on the platform.
- 25 47. The method of claim 46 further comprises the step of providing a feedback mechanism associated with the intelligent entity spot market, the feedback mechanism being configured or configurable to record each and every step in the problem solving process, and to create a vector track record of a performance of all the intelligent entities on the task or the sub-task.
48. The method of claim 47, wherein the vector track record is implemented by way of blockchain
30 technology to allow for precise reputations that are analyzed by an AI agent or system and converted into estimates of a value for each of the intelligent entities for any of the tasks.
49. The method of claim 39 further comprises the step of compensating the intelligent entity when exiting the online Ad advertisement unit and stops participating in the problem solving process.

50. The method of claim 39 further comprises the step of crediting the intelligent entities if the task or the sub-task is solved, based on an amount of contribution by the participating intelligent entities.

51. The method of claim 39 further comprises the step of improving the interactive online advertisement unit based on a feedback loop including on any one of or any combination of
5 reputational metrics, and metrics related to the problem solving process or knowledge captured.

52. The method of claim 2, wherein the step of creating a self-aware operation for the intelligent entities by adding a dimension of self-awareness and increased autonomy to the intelligent entities further comprises the step of:

equipping any one of or any combination of the intelligent entities with one or more components
10 configured or configurable to operate with characteristics of a spotlight of attention model;
setting dynamic parameters for working memory of the intelligent entity, respectively, that
corresponds to cognitive resource limits;
providing a dimension of categorization for events in the working memory that relates to self or
non-self;
15 categorizing each of the events, as the events are encountered by the intelligent entity,
respectively, with respect to categories that the intelligent entity wishes to be aware of; and
constructing a model of a state of awareness for the intelligent entity, the model consisting of a
total of the events that are active in the working memory based on the parameters, for each
of the categories of awareness, including a current self and environmental state of
20 awareness.

53. The method of claim 52, wherein the step of equipping the intelligent entities with the components includes any one of or any combination of:

an input system configured for sensory and non-sensory cognitive input or perceptual inputs and
self-generated concepts;
25 an attention mechanism configured or configurable to focus computational resources of the
intelligent entities on specific stimuli that are relevant at any given time;
pattern recognition algorithms configured or configurable to compare the sensory and non-
sensory cognitive input or the perceptual inputs with the working memory to recognize
objects and events, and identify which elements within the sensory input or the working
30 memory are likely to be relevant to the task of the intelligent entities, the pattern recognition
algorithms are further configured or configurable to categorize and store information in a
structured manner for future retrieval;
memory systems configured or configurable to support the working memory, short-term
memory, and long term memory capabilities;

categorization capabilities configured or configurable to process the sensory and non-sensory cognitive input or the perceptual inputs and to categorize the inputs into various classes including perceptual events, cognitive events, interactions; and self-referential events, and concept formation capabilities that enable the intelligent entities to
5 form new human-understandable concepts.

54. The method of claim 52 further comprising the step of monitoring and updating the categories of awareness of the intelligent entities.

55. The method of claim 54, wherein the step of monitoring and updating the categories of awareness further comprises the steps of:

- 10 retrieving, by the intelligent entities, existing categories of awareness;
- maintaining an awareness in parallel with other problem solving operations of the task provided to the intelligent entities;
- monitoring and updating continuously the categories of awareness of the intelligent entities in real-time to change the state of awareness of the intelligent entities; and
- 15 providing a feedback loop to refine the categories of awareness.

56. The method of claim 55, wherein the step of monitoring and updating continuously the categories of awareness includes the steps of:

- using an attention mechanism configured or configurable to direct attention of the intelligent entities periodically from the problem solving task to updating the state of awareness;
- 20 enabling attention interrupts that are configured or configurable to shift attention immediately from the problem solving task if any external perception or internally self-generated concept from an input system detects one or more of the events that matches of list of events constituting intentional interrupts; and
- updating the state of awareness when the attention is directed.

25 57. The method of claim 52 further comprises the step of resolving a conflict in behavior of any one of the intelligent entities based on differing identities and self-concepts.

58. The method of claim 57 further comprises the step of providing ethical reasoning and consequence prediction that comprises the steps of:

- identifying a conflict between a behavioral directives of two or more of the active identities, the
- 30 recognizing of the conflict utilizes a voting method from the intelligent entities;
- gathering information that collects relevant data about the situation, including the potential consequences of the different actions, relevant ethical principles, and human safety considerations;

providing simulation options that utilize a virtual environment to simulate potential actions and consequences under the recognized conflict of each of the active identities;
evaluating and prioritizing that analyzes predicted outcomes of each of the actions, prioritizing actions that minimize harm to humans and align with the ethical principles; and
5 selecting and implementing the action that best resolves the conflict while adhering to ethical guidelines and minimizing risk to humans, documenting a reasoning process for future reference and learning.

59. The method of claim 57 further comprises the step of providing hierarchical override with justification that comprises the steps of:

- 10 identifying a conflict between behavioral directives of two or more of the active identities;
providing a reference hierarchy that consults an established hierarchy of the identities, where human safety and well-being attributes holds a highest priority;
providing a means to activate override where the identities higher in the hierarchy takes precedence;
- 15 providing justification and transparency that documents the conflict, a decision-making process, and a justification for a chosen action based on the hierarchy and ethical principles; and
providing learning and adaptation that learns from experience, and refines an understanding of the conflicting identities and adjusting the hierarchy or the behavioral protocols to prevent similar conflicts in the future.

20 60. The method of claim 57 further comprises the step of providing external arbitration and input from the intelligent entities that comprises the steps of:

- recognizing intractable conflict that identifies a conflict that cannot be resolved independently due to a complexity of a situation or an equally weighted importance of conflicting identities;
- 25 seeking external input that requests guidance from external intelligent entities or a designated ethics committee, and providing all relevant information about the conflict, potential actions, and predicted consequences;
providing collaborative deliberation wherein the intelligent entities and intelligent entity collaborators engage in a discussion, considering ethical principles, human values, and potential consequences of different actions;
- 30 providing joint decision-making based on the collaborative deliberation, a course of action is chosen that aligns with both core principles and human ethical considerations; and

providing documentation and learning that documents the conflict, a resolution process, and a rationale behind a final decision, for improving an ability to handle similar conflicts in the future.

61. The method of claim 57 further comprises the step of providing identity negotiation and compromise that comprises the steps of:

5 identifying shared goals that analyzes conflicting identities and seeks to identify any underlying shared goals or values;
exploring alternative actions that potentially satisfy core principles of both conflicting identities;
evaluating compromise options that assess potential consequences of each compromise option,
10 prioritize solutions that minimize harm to humans and uphold key ethical principles;
select and implementing compromise that chooses the compromise that best balances needs of the conflicting identities while prioritizing human safety and well-being; and
monitoring and adapting that observes outcomes of the chosen action and makes adjustments as needed to ensure that the compromise remains effective and aligned with ethical
15 considerations, and that learns from the experience, refining its understanding of the conflicting identities and adjusting a hierarchy or behavioral protocols to prevent similar conflicts in the future.

62. The method of claim 57 further comprises the step of providing temporary identity suspension that comprises the steps of:

20 identifying destructive conflict that recognizes a conflict between two or more of the identities that, if acted upon, could lead to actions that directly harm humans or violate fundamental ethical principles;
isolating the conflicting identity and temporarily suspending behavioral protocols of the identity that poses a most direct threat to human safety or ethical integrity;
25 proceeding with an alternative identity that proceeds with a guidance of one or more of remaining active identities, ensuring actions align with human safety and well-being;
providing reflection and reintegration, during the temporary suspension, that reflects on reasons behind the conflict and explores potential modifications to behavioral protocols of the suspended identity to prevent future conflicts; and
30 providing gradual reintroduction that reintroduces the suspended identity with updated protocols, ensuring its alignment with the priority of human safety and ethical behavior.

63. The method of claim 52, wherein the model of the state of awareness is a foundational model of awareness for any one of the intelligent entities, foundational model of awareness comprises the steps of:

- logging into a website by any one of the intelligent entities;
selecting one or more training algorithms for the foundational model from a set of training techniques found on the website or from machine learning algorithms;
selecting one or more training datasets that reflects any one of or any combination of expertise,
5 knowledge, ethical preferences, values and personality of the human user;
training the foundational model using the selected training algorithms and the selected training datasets;
training the foundational model to explicitly operate a spotlight of attention;
recording, during all interactions, what is within the spotlight of attention, and identifying in the
10 record whether each item that is attended to constitutes “self” or “not-self”;
interacting with and instructing the trained foundational model to form a self-concept and identity that is reflected in the training materials;
instructing the trained foundational model to continuously monitor one or more inputs to the trained foundational model for elements that change a sense of self-awareness of the
15 intelligent entities, and to maintain and auditable record of how a concept of self-awareness of the intelligent entities is changing based on the inputs as well as boundaries that currently define a dynamically changing sense of self;
refining and improving an output of the trained foundational model based on dialog and interaction with the trained foundational model until the trained foundational model behaves
20 like the human user so that the trained foundational model passes a Turing Test involving other human users who know the human user; and
subjecting the trained foundational model to the Turing Test, when the human user is satisfied with a progress of the intelligent entities.
64. The method of claim 63, wherein the logging into the website is performed from a social media
25 platform.
65. The method of claim 63, wherein the Turing Test further comprises the steps of:
identifying the other intelligent entities who know the intelligent entity, or that are determined to be helpful in discriminating between humans and AIs;
interacting the identified other intelligent entities with the foundational model and with the
30 intelligent entity utilizing a questionnaire provided to the identified other intelligent entities, the questionnaire including questions require an identity or sense of self to answer;
predicting by the identified intelligent entities which of the intelligent entity was a human and which was the foundational model, and providing a confidence estimate for the prediction;

performing a statistical analysis on the predictions of the identified intelligent entities and on ratings of the identified intelligent entities, the statistical analysis is configured or configurable to determine whether the predictions were able to identify the intelligent entity as a human; and

- 5 repeating the step of training or tuning the foundational model, on condition that the foundational model is distinguishable from the intelligent entity or within a preset level of statistical significance, with adjustments to any one of or any combination of the machine learning algorithms, and the training datasets to change or tune the foundational model until a behavior of the foundational model becomes indistinguishable, as measured by the preset
10 level of statistical significance, from that of the intelligent entity or the foundational model needs to be modified further before additional training or tuning.

66. The method of claim 52 further comprises the steps of:

forming new identities and self-concepts of any one of the intelligent entities dynamically; and determining which of the identities and self-concepts is active at any given moment.

- 15 67. The method of claim 66 further comprises the step of providing a hierarchical identity structure with ethical override that comprises the steps of:

establishing a hierarchical structure configured or configurable to organize the identities in a hierarchy with human safety and well-being attributes at an apex of the hierarchy;

- 20 identity activation configured or configurable to use contextual cues and current goals to determine a most relevant identity for a situation of the AI agent or system;

resolving conflict by dictating a behavior of the intelligent entities based on the hierarchy dictates;

- 25 providing an ethical reasoning engine that continuously evaluates potential consequences of actions of the intelligent entities based on all the active identities; and

performing learning and adaptation that learns from experiences and feedback, and refines one or more of the identities within the hierarchy.

68. The method of claim 66 further comprises the step of providing identity-specific behavioral protocols that comprises the steps of:

- 30 providing protocol development including for each of the active identities, a set of behavioral protocols is defined and refined by way of interactions with other intelligent entities, wherein the protocols outline acceptable actions, decision-making processes, and limitations based on principles of the active identities, respectively;

- providing identity recognition that analyzes a current situation, including information that is within a spotlight of attention to identify a relevant identity and activate corresponding behavioral protocols of that relevant identity;
- providing action selection, within the active protocols, that selects actions that are most likely to achieve the task while adhering to principles of the active identities and prioritizing human safety;
- providing feedback and refinement where outcomes of actions are continuously evaluated, and the protocols are adjusted to improve future performance and alignment with a set of core values of each of the active identities; and
- providing external review by periodically reviewing the protocols for each of the identities by other intelligent entities.
69. The method of claim 66 further comprises the step of providing identity simulation and consequence prediction that comprises the steps of:
- creating a simulation environment that includes a secure virtual environment where different scenarios and potential actions under each of the active identities is simulated;
- providing consequence prediction that is configured or configurable to estimate a likely consequences of actions within the simulation, focusing on potential impacts on human safety and well-being;
- providing evaluation and selection that evaluates the consequence prediction and selects an action that best aligns with principles of the active identities while minimizing risk to human safety;
- providing real-world implementation and monitoring that implements the selected action in the real world utilizing the network, and closely monitors results of the selected action by comparing to the predicted outcomes; and
- providing continuous learning that incorporates the results of each of the simulations and the results of the selected action in the real world action into a knowledge base, and refines an understanding of each of the identities, and improves an ability to predict consequences.
70. The method of claim 66 further comprises the step of providing identity-based moral dilemma training that comprises the steps of:
- providing a scenario database that includes scenarios and moral dilemmas covering various situations relevant to the identities;
- providing dilemma presentation that presents the intelligent entities with the scenarios and moral dilemmas, and tasks them with analyzing the scenarios and moral dilemmas from a perspective of the relevant identity;

- providing ethical reasoning and justification that applies principles and values of the active identity to reason through the scenarios and moral dilemmas, and that generates solutions and justifications to the scenarios and moral dilemmas;
- providing intelligent entity evaluation and feedback that reviews reasoning and the solutions by the intelligent entities, and provides feedback on alignment of the solutions with human values and safety priorities; and
- providing iterative learning and improvement that refines ethical reasoning skills and an ability to make sound judgments aligned with human safety within the context of each of the identities by repeated exposure to the scenarios and moral dilemmas and the feedback.
71. The method of claim 66 further comprises the step of providing collaborative identity development with input from the intelligent entities that comprises the steps of:
- providing intelligent entity interaction that engages in regular interactions and dialogues with diverse groups of other intelligent entities representing various cultures, backgrounds, and belief systems;
- providing identity exploration, through the interactions, to gain an understanding of human and other intelligent entity perspectives on various identities and their associated values, principles, and behaviors;
- providing collaborative refinement that collaborators work together to refine definitions and behavioral protocols for each of the identities, ensuring they remain consistent with human values and ethical principles;
- providing human-in-the-loop decision making that seeks input and guidance from human collaborators, or an intelligent entity representative certified and approved by humans for critical decisions or situations; and
- providing continuous co-evolution that utilizes ongoing interactions and feedback from humans or humans' intelligent entity representatives.
72. The method of claim 2, wherein the step of searching through any one of or any combination of a collective network of AIs, a collective network of AAAs, a collective network of AGIs, and a collective network of PIs for new information that is different to a current information further comprises the steps of:
- participating multiple of the intelligent entities in the collective AGI network;
- providing payment by one or more of the intelligent entities for problem solving or other cognitive work on the task on the AGI network;

reserving a portion of the payment to cover operating costs for the one or more of the intelligent entities that is providing the problem solving or other cognitive work on the task, including a reserve that is allocated to expand the AGI network; and

5 determining when some of the AGI network is not engaged in a problems solving operation, then the intelligent entities are utilized to expand and extend the AGI network following the universal problem solving architecture and framework.

73. The method of claim 72, wherein the step to expand and extend the AGI network further comprises the steps of:

10 setting a default task to expand the AGI network; and
running safety and ethics checks each time a task or sub-task is set and also before each potential action is taken as long as spare capacity and resources of the intelligent entity exist to work on the task or sub-task.

74. The method of claim 73, wherein the step of running safety and ethics checks further comprises the steps of:

15 recruiting one or more of the intelligent entities to solve the task; and
representing, by the recruited intelligent entities, the task as one of achieving a series of sub-tasks.

75. The method of claim 74, wherein the step of representing the task as a series of sub-tasks further comprises the steps of:

20 increasing an intelligence of the intelligent entities on the collective network of AGIs;
recruiting additional human users to t the collective network of AGIs;
using the increased intelligence intelligent entities and additional human users to determine bottlenecks to increase expansion of the collective network of AGIs;
prioritizing the bottlenecks such that the ones that lead to a greatest benefit in terms of network
25 expansion are solved first; and
applying one or more problem solving techniques to solving each of the bottlenecks and expanding the collective network of AGIs.

76. The method of claim 75 further comprises the steps of:

30 repeating the steps of prioritizing the bottlenecks and applying one or more problem solving techniques until:
diminishing returns occur and then switching to the step of increasing the intelligence of the intelligent entities on the AGI network as opposed to increasing a scope of the AGI network; or

spare resources are exhausted and then pausing the solving of the bottlenecks awaiting additional resources from solving other tasks.

77. A method for Planetary Intelligence (PI) with human-aligned behavior utilizing a collective network of intelligent entities selected from the group consisting of any combination of a user computer system operated by a human user, an Artificial Intelligent (AI) agent or system, an Advanced Autonomous Artificial Intelligences (AAAI) agent or system, an Artificial General Intelligent (AGI) agent or system, a SuperIntelligent (SI) agent or system and a Personalized SuperIntelligence (PSI) agent or system, the method comprising the steps of:

- establishing a collective network of the intelligent entities;
- 10 utilizing online advertising technology for increasing an intelligence of the intelligent entities;
- implementing a universal problem solving architecture and framework on a task or a sub-task of the task provided to any one of the intelligent entities to collaborate and create higher levels of intelligence;
- utilizing a spot market to solicit for attention of any one of the intelligent entities;
- 15 utilizing a human-centered input configured or configurable for obtaining values and ethics information from multiple human users and then using the values and ethics information to customize the AI agent or system to create the AAAI agent or system;
- creating a collective network of PSIs by combining an intelligence of the AAAI agent or system that forms a collective network of AAAIs;
- 20 creating a collective network of AGIs by combined intelligence of any one of or any combination of the collective network of AAAIs, a collective network of AI agents or systems, the collective network of PSIs, and a collective network of user computer systems;
- creating a self-aware operation for each of the AGI agents or systems in the collective network of the AGIs by adding a dimension of self-awareness and increased autonomy to the AGI
- 25 agents or systems, and creating an ability to assume multiple identities of the AGI agents or systems to handle tasks that arise in parallel with other AGI agents or systems; and
- creating the PI agent or system comprising the collective network of AGIs that work together toward a solution to the task or the sub-task.

78. The method of claim 77, wherein the spot market includes:
- 30 a means for the intelligent entities with attention to sell to access a marketplace and specify seller information for sale and an ask price for the information;
 - a means for buyers of intelligent entity attention to access the marketplace and specify buyer information the buyer is willing to buy and a bid price for buying the buyer information;

a market mechanism for queuing the bid price and the ask price, including categories of the buyer and seller information, wherein each category has a market in the marketplace; the market mechanism is configured or configurable to make the market in each category by market makers; and

- 5 the market mechanism is configured or configurable to match bid prices and ask prices, and a transaction occurs that is binding on the buyer and the seller of the information.

79. The method of claim 77, wherein the spot market includes the steps of:

- buyers and sellers of intelligent entity attention and expertise register on a platform, the buyers providing details about buyer interests or expertise, the sellers providing details about sellers
10 interests or expertise;

listing by the sellers available time slots and expertise areas, and listing by the buyers needs and time slots the buyers is interested in;

utilizing, by the platform, an algorithm to dynamically price intelligent entity attention and expertise based on supply, demand, and user ratings;

- 15 matching one or more of the buyers and one or more of the sellers based on requirements, availability, and price;

enabling transactions where the buyers pay for the seller time slots, and wherein the platform takes a commission; and

- 20 providing feedback where after each session, the buyers and the sellers rate each other, influencing future pricing and matching.

80. The method of claim 77, wherein the spot market includes the steps of:

buyers and sellers of intelligent entity attention and expertise register on a platform, the buyers providing details about buyer interests or expertise, the sellers providing details about sellers
interests or expertise;

- 25 creating an auction by the sellers for seller time slots;

placing one or more buyer bids by a buyer on time slots and expertise the buyer requires;

closing the auction at a predetermined time or when the seller accepts a buyer bid;

paying the seller by the buyer of a winning bid, and the seller provides the attention or expertise
based on the time slot, wherein the platform mediates an exchange and secures payment; and

- 30 providing feedback where after each auction, the buyers and the sellers rate each other, affecting future auctions and visibility on the platform.

81. The method of claim 80 further comprises the step of providing a feedback mechanism associated with the intelligent entity spot market, the feedback mechanism being configured or

configurable to record each and every step in the problem solving process, and to create a vector track record of a performance of all the intelligent entities on the task or the sub-task.

82. The method of claim 81, wherein the vector track record is implemented by way of blockchain technology to allow for precise reputations that are analyzed by an AI agent or system and converted
5 into estimates of a value for each of the intelligent entities for any of the tasks.

83. The method of claim 77, wherein the step of utilizing online advertising technology for increasing an intelligence of the intelligent entities further comprises the steps of:

populating an online advertisement unit including information associated with the task;
providing the online advertisement unit to one or more of the intelligent entities utilizing the
10 collective network;
receiving one or more inputs from the intelligent entities participating in the online advertisement unit, and the inputs being associated with any one of or any combination of solving the task, solving a sub-task of the task, and advancing progress on the task;
communicating the inputs to the any one of or any combination of the intelligent entities;
15 utilizing the inputs in a universal problem solving method of the universal problem solving architecture and framework to generate a solution to the task or the sub-task; and
training any one of or any combination of the intelligent entities with results from the universal problem solving method that are based on successful or unsuccessful solution attempts to solve the task or the sub-tasks, thereby increasing an intelligence of the intelligent entity,
20 respectively.

84. The method of claim 83 further comprises the step of receiving advertisement specifications from a client to be used in the populating or targeting of the online advertisement unit in combination with information associated with the task.

85. The method of claim 84 further comprises the step of billing the client based on cost per
25 thousand impressions or clickthrough rate metrics.

86. The method of claim 84, wherein the advertisement specifications include any one of or any combination of advertisement content, demographic information, location restrictions for the online advertisement unit, advertisement budget, and metrics for determining a successful solving of the task or a sub-task.

30 87. The method of claim 77 further comprises the step of executing a safety and ethics check at any one of or any combination of each time the task or the sub-task is set, before each potential action is taken on the task or the sub-task on, or on the solution.

88. The method of claim 87, wherein the step of implementing the universal problem solving architecture and framework further comprises the steps of:

acquiring information associated with the task from any one of or any combination of the intelligent entities;

identifying one or more of the intelligent entities that have one or more attributes related to one or more request criteria of the task;

- 5 implementing by each of the identified intelligent entities the universal problem solving architecture and framework on the task to create a completion solution; and
providing the completion solution to any one of or any combination of the intelligent entities for final acceptance.

89. The method of claim 88 further comprises the steps of:

- 10 comparing any one of or any combination of the task, and the solution against prohibited attributes, and assigning an ethics attribute to one of or any combination of the task, and the solution based on any one of or any combination of a result of the comparison, and an ethics criteria;
implementing, based on the result of the comparison, the universal problem solving architecture
15 and framework on the task to create the solution and creating an AGI; and
providing the results of the comparison and the solution to any one of the intelligent entities and additional intelligent entities on the collective network.

90. The method of claim 89 further comprises the step of recording one or more problem solving activities from each of the intelligent entities in an auditable record, and comparing the problem
20 solving activities with a successful or unsuccessful progress towards the solution of the task, and determining which of the problem solving activities to keep active.

91. The method of claim 89 further comprises the step of learning by the intelligent entities a procedural learning process of the universal problem solving architecture and framework, wherein the intelligent entities provide information to the procedural learning process for creation of the
25 AGI.

92. The method of claim 89, wherein the ethics criteria include a confidence level threshold for the goal so that the ethics attribute is determined as any one of an unsafe goal, an unethical goal, a safe goal, and an ethical goal.

93. The method of claim 92, wherein the confidence level threshold is further utilized to determine
30 if a sequence of individually safe goals is unsafe or unethical when considered cumulatively.

94. The method of claim 92, wherein the confidence level threshold is utilized to determine whether a violation occurred that reflects a predictive evaluation if the goal is to violate the ethics criteria.

95. The method of claim 88 further comprising the step of cloning any one of the intelligent entities for deployment of multiple copies thereof to assist in any one of or any combination of creating of the solution, or providing training data to any one of the intelligent entities.

96. The method of claim 95 further comprising the step of estimating a worth of the cloned intelligent entities utilizing a network effect value including the number of cloned intelligent entities available on the collective network.

97. The method of claim 96 further comprising the step of utilizing the estimated worth for determining pricing decisions for problem solving services offered by the cloned intelligent entities on any one of the social media platforms or through an additional intelligent entity.

98. The method of claim 97 further comprising the step of monetizing the cloned intelligent entities for each utilization of the cloned intelligent entities on the social media platforms or the additional intelligent entity.

99. The method of claim 97 further comprising the step of allowing access to the cloned intelligent entities by any one of the social media platforms so that the social media platforms can receive the solution to the task or using the training data for an AI system of the social media platform.

100. The method of claim 77 further comprises the step of increasing a knowledge of the intelligent entities by incorporating new sources of informational datasets that differs from a current informational dataset of the intelligent entity.

101. The method of claim 100, wherein the step of increasing a knowledge of the intelligent entities by incorporating new sources of informational datasets that differs from a current informational dataset of the intelligent entity further comprises the steps of:

searching for one or more potential informational datasets from one or more sources, the potential informational datasets being related to a knowledge dataset of the intelligent entity, respectively;

determining a difference of the potential informational datasets by utilizing a difference attribute of the potential informational datasets with regard to one or more factors; and learning by utilizing the potential informational datasets based on the difference attribute of the potential informational datasets.

102. The method of claim 101 further comprises the steps of:

sampling subsets of the potential informational datasets and calculating a goal-relevancy attribute to identify one or more of the sampled subsets that have a highest goal-relevancy; estimating a Shannon Entropy on the one or more sampled subsets; calculating a Kaplan Information Theoretical (KIT) relevance utilizing a product of the Shannon Entropy and the goal-relevancy attribute of each of the subsets;

grouping the subsets based on the KIT relevance to determine a first approximation of an optimal grouping of the subsets including a prioritized grouping of the potential informational datasets; and

5 providing the prioritized grouping of the potential informational datasets to the intelligent entities for learning by the any one of the intelligent entities.

103. The method of claim 77, wherein an operational aspect of the PI is self-funded by compensation received by any one of or any combination of the online advertising technology, and the spot market.

104. The method of claim 77, wherein the human user participates in problem solving of the task or the sub-task utilizing the user computer system and the universal problem solving architecture and framework, thereby creating a human-centered input.

105. The method of claim 104, wherein the human-centered input further comprises the steps of:

15 matching one or more human workers from a data source including a list of human problem solvers to the task based on a task criteria;

translating any part of the task into an unambiguous language utilizable in the universal problem solving architecture and framework including a decision tree;

separating the task into sub-tasks;

delegating each of the sub-tasks to one or more of the matched human workers so that work on each of the sub- tasks proceeds independently from each other and parallel with each other;

20 utilizing the universal problem solving architecture and framework in a problem solving process on the sub-tasks, respectively, to create one or more sub-solutions;

receiving the sub-solutions from each of the matched human workers for the sub-tasks delegated thereto;

combining the sub-solutions into an overall solution to the task;

25 directing any one of or any combination of a new human worker from the data source and one or more of the matched human workers to parts of the decision tree where work is required;

compensating the matched human workers for the sub-solutions;

providing any one of or any combination of the sub-solutions and the overall solution to any one of or any combination of the intelligent entities;

30 allowing any one of or any combination of the intelligent entities to accept the overall solution, reject the overall solution, and provide feedback to any one of the matched human workers on any one of the sub-solutions; and

assigning a reputation attribute to any one of or any combination of the human workers and the intelligent entities.

106. The method of claim 105, wherein the decision tree is maintained in blockchain or Ethereum logs.

107. The method of claim 105, wherein the reputation attribute includes metrics on any one of or any combination of a time to the sub-solutions, a difficulty value of the task, short and long-term
5 user satisfaction with the sub-solutions, a number of times any one of the sub-solutions has been re-used on the collective network, a rating by other human workers, a responsiveness value of the human workers, and a reliability value of the human workers.

108. The method of claim 105 further comprising the step of using the reputation attribute in the matching of the human workers to the task using an algorithm to the delegation of the sub-task.

109. The method of claim 105 further comprising the step of soliciting, at predetermined intervals
10 after the overall solution or the sub-solutions are provided to any one of the intelligent entities, feedback by way of a survey for user satisfaction information to obtain satisfaction metrics that are used to update the reputation attribute of one or more of the human workers or the intelligent entities, respectively.