SUPERINTELLIGENCE
AN iQ COMPANY

ABSTRACT & SUMMARY
SUPERINTELLIGENCE DESIGN WHITE PAPER #10: PLANETARY INTELLIGENCE

by Dr. Craig A. Kaplan
May 2025

ABSTRACT

Planetary Intelligence (PI) requires the cooperation of multiple Artificial General Intelligences (AGIs). These AGIs collaborate over a self-extending, global, problem-solving network. Each AGI comprises a network of (human and AI) intelligent entities. Each AI entity can be customized and personalized with specific human values and knowledge.

Alignment results from many AI entities combining their human-centered values democratically, using representative and statistically valid methods. Safety is designed into the system and scales as AGIs and PI increase intelligence and speed.

Our modular, scalable design of PI integrates more than 100 novel inventive systems and methods. Specific inventions include: a universal problem-solving architecture and methods; new methods for AI learning and customization; methods for integrating intelligent entities; catalysts for increasing intelligence; superior monetization methods; attentional systems enabling awareness, and self-awareness; methods for ethical conflict resolution; and methods for maximizing human-alignment and the safety of AI, AGI, and PI systems.

SUMMARY

White Paper #10 describes a new architecture and method for creating a global, superintelligent Artificial General Intelligence (AGI) system called Planetary Intelligence (PI). The author claims that PI is the next logical step in the evolution of Artificial Intelligence. It can be created by networking together many AGI systems designed with human-aligned ethics and safety features. The author emphasizes the importance of the PI architecture for achieving safe AI and believes that the methods disclosed in this PPA represent the fastest and safest path to the development of PI. White Paper #10 is the culmination of nine previous white papers that describe various aspects of AI, AGI, and SuperIntelligent systems. White Paper #10 shows how

dozens of inventions and designs of smaller components can be integrated into an intelligence of global scale – a Planetary Intelligence network.

Novel Features of the White Paper

- **A unique approach to creating and managing the global network of AGI systems that comprise PI.** The author describes a "collective intelligence" approach in which intelligent entities collaborate using a common problem–solving framework, including humans, AI agents, and AI systems.
- **The white paper presents the concept of "spot markets" to acquire the expertise of human or non-human intelligent entities and the idea of "reputation" for guiding the acquisition and allocation of expert resources.** This approach is intended to accelerate the progress of PI by facilitating access to the best information and knowledge and allowing PI to monetize its resources.
- **A sophisticated system for ensuring that AI systems are aligned with human values and safe.** This system includes a variety of methods for identifying, eliciting, and incorporating human values into the design and training of AI systems, as well as for resolving conflicts between different value systems. The system also includes a robust safety framework that incorporates mechanisms for detecting and preventing potential threats to human safety.
- **A method for extending the scope of PI by developing self-extending networks of AGI systems.** The author explains that AI systems have a natural tendency to expand their intelligence and to integrate with other systems, and that this tendency can be leveraged to create PI.

Detailed Description of Each Section of the White Paper

Introduction: This section introduces the concept of PI and explains the rationale for developing PI as a global, superintelligent system. The author also describes the role of earlier white papers in guiding the development of an overall PI system.

Background Art: This section cites several previous papers relevant to the development of PI and incorporates these by reference.

Stakes for Humanity: This section describes PI's potential risks and benefits to humanity. The author provides data on the expected number of deaths from AI and compares the estimated deaths to those in major wars in the past 200 years. The author argues that the potential dangers of AI are much greater than the dangers of war, and that humanity must act quickly to address these dangers.

Some Features of the Design That Reduce Risk of Extinction by AI: This section highlights the safety features of the design, including the human-centered design, collective intelligence and diversity of perspectives, transparency and auditability, continuous learning and adaptation, safety mechanisms and safeguards, and the alignment problem.

Some Significant Remaining Risks to Humanity: This section details several risks that remain despite best efforts to maximize the safety of the design for PI. These risks include unforeseen consequences of emergent properties, evolution of values and ethical frameworks, concentration of power and influence, vulnerability to cyberattacks and system failures, and existential risks from self-aware AI.

OVERVIEW OF THE DESIGN OF PLANETARY INTELLIGENCE: This section provides an overview of the PI system, including a list of definitions used throughout the white paper.

Definitions: This section provides definitions for key terms, including:

- Artificial Intelligence (AI)
- Artificial General Intelligence (AGI)
- Advanced Autonomous Artificial Intelligence (AAAI)
- AI Ethics
- Alignment Problem
- Awareness
- Base AI
- Collective Intelligence (CI)
- Human Ethics
- Intelligent Entities
- Inter-Planetary Intelligence (IPI)
- Large Language Model (LLM)
- Machine Learning (ML)
- Narrow AI
- Personalized SuperIntelligence (PSI)
- Planetary Intelligence (PI)
- Prohibited Attributes
- Safety
- Safety Feature
- Self-Awareness
- Self-Concept
- SuperIntelligence (SI)

- Training/Tuning/Customization
- Weights/Weights of the Network

Key Dimensions of a PI System: This section discusses the key dimensions of a PI system, including modular architecture, universal problem-solving capabilities, human-centered design, knowledge and expertise integration, personalization and customization, growth of intelligence, safety, and acquiring and aligning values with human values.

Specific Inventive Systems & Methods for Implementing PI: This section summarizes the specific inventive systems and methods that may be used to implement PI, organized by the dimensions discussed in Section 2.2. This section includes an extensive list of figures illustrating key aspects of the inventive methods and systems.

Summary of Previous Inventive Methods and Systems: This section summarizes the key inventive methods and systems from prior white papers, relevant to White Paper #10.

Implementation of a Planetary Intelligence: This section describes implementing a PI system using the disclosed systems and methods. This section is divided into four subsections:

High-level Description of PI System: This section provides a high-level description of the PI system. The author explains that PI is a network of AGIs that are networks of intelligent entities (including humans, AI agents, and AI systems).

Principal Components and Categories of Supporting Systems and Methods for Exemplary PI Architecture: This section provides a more detailed explanation and illustration of the PI architecture. The author also explains the role of online advertising technology in funding the PI system.

Detailed Mapping of Inventive Systems and Methods to Exemplary PI Architecture: This section thoroughly maps specific inventive systems and methods from prior white papers to the exemplary PI architecture described in White Paper #10.

Exemplary Specific PI Implementation Using Subset of Systems and Methods: This section provides an exemplary implementation of a PI system, based on the assumption that the author is the CEO of a large technology company like META. The author describes META's steps to implement a PI system using the systems and methods disclosed.

Concluding Remarks: White Paper #10 discussed the importance of safety and ethics in developing PI. The author argues that human values must be incorporated into the design of PI systems and that PI systems must be designed to behave ethically. The author also emphasizes

the need for humans to guide the development of PI, and he warns that if humans do not act quickly to address the dangers of PI, then humanity could face an existential threat.

Diagrams

A separate file contains more than 100 diagrams illustrating the various components of PI systems, each described briefly in White Paper #10 and in more detail in the previous white papers. The titles of the various Figures, along with a brief description of what the Figures describe, are a quick way to gain a sense of some of the key inventions involved in the overall design of a PI and how these inventions fit together.

- **FIG. 1-1**: This diagram illustrates the "Universal Problem-Solving Architecture" used to coordinate the activities of the intelligent entities comprising PI.
- **FIG. 1-2**: This diagram illustrates the network concept used to connect multiple intelligent entities.
- **FIG. 1-3**: This diagram illustrates a decision-tree structure for problem solving.
- **FIG. 1-4**: This diagram illustrates the WorldThink Architecture, which coordinates the activities of intelligent entities with specific expertise or skill.
- **FIG. 1-5**: This diagram illustrates creating, training, and customizing an AAAI.
- **FIG. 1-6**: This diagram describes the Universal Problem-Solving Architecture.
- **FIG. 1-7**: This diagram illustrates how problem-solving proceeds in a collective intelligence network.
- **FIG. 1-8**: This diagram illustrates a process of procedural learning.
- **FIG. 1-9**: This diagram illustrates a solution learning system.
- **FIG. 1-10**: This diagram illustrates a network of networks.
- **FIG. 2-1**: This diagram illustrates serial problem solving.
- **FIG. 2-2**: This diagram illustrates parallel problem solving.
- **FIG. 2-3**: This diagram illustrates cloning.
- **FIG. 2-4**: This diagram illustrates problem solving on a network.
- **FIG. 2-5**: This diagram illustrates the Universal Problem-Solving Framework.
- **FIG. 2-6**: This diagram illustrates the hierarchical tree construct.
- **FIG. 2-7**: This diagram illustrates a WorldThink Tree.
- **FIG. 2-8**: This diagram illustrates a method for embedding safety checks into the problem-solving architecture.
- **FIG. 2-9**: This diagram illustrates a method for embedding safety checks into the problem-solving architecture.
- **FIG. 2-10**: This diagram illustrates a method for recording and updating context for intelligent entities.

- **FIG. 2-11**: This diagram illustrates a method for translating natural language into a formal problem-solving language.
- **FIG. 3-1**: This diagram illustrates a general process for customizing AAAIs.
- **FIG. 3-2**: This diagram illustrates methods for eliciting human-aligned ethical preferences.
- **FIG. 3-3**: This diagram illustrates a method for automatically generating questionnaires.
- **FIG. 3-4**: This diagram illustrates a general method for identifying values.
- **FIG. 3-5**: This diagram illustrates a method for customizing AI.
- **FIG. 3-6**: This diagram illustrates a general method for training AI.
- **FIG. 3-7**: This diagram illustrates a general method for creating scalable, ethical AGI using the customized AIs.
- **FIG. 3-8**: This diagram illustrates a method for adding a reputational component to AI systems.
- **FIG. 3-9**: This diagram illustrates customization of AI systems.
- **FIG. 3-10**: This diagram illustrates additional customization of AI systems.
- **FIG. 4-1**: This diagram illustrates a voting method for combining ethical and other information from multiple customized AI agents.
- **FIG. 4-2**: This diagram illustrates a method for using problem solving to refine values once ethical or other information from customized AI agents has been combined.
- **FIG. 4-3**: This diagram illustrates a method for scalable AGI that includes steps for combining information from weight matrices.
- **FIG. 4-4**: This diagram illustrates a method for scalable AGI that includes steps for combining information, testing, and monitoring the combination results.
- **FIG. 4-5**: This diagram illustrates a consensus method for preventing hallucination.
- **FIG. 4-6**: This diagram illustrates methods for using knowledge modules and collections of agents to customize AI.
- **FIG. 4-7**: This diagram illustrates a method for combining information from multiple intelligent entities.
- **FIG. 5-1**: This diagram illustrates a general process for implementing PSI.
- **FIG. 5-2**: This diagram illustrates how PSI, AGI, and PI can leverage their abilities to increase intelligence.
- **FIG. 5-3**: This diagram illustrates a critical community-based safety mechanism in which PSI can serve as a check on other PSIs in a network.
- **FIG. 5-4**: This diagram illustrates methods for recording the actions of AI, AAAI, PSI, AGI, and PI on their respective networks.
- **FIG. 5-5**: This diagram illustrates methods and checks of cognitive activity.
- **FIG. 5-6**: This diagram illustrates a method using competition and evolution to increase the intelligence of an AI system.
- **FIG. 5-7**: This diagram illustrates various characteristics of an intelligent network.

- **FIG. 5-8**: This diagram illustrates problem-solving tasks that can be useful in increasing the intelligence of AI systems.
- **FIG. 5-9**: This diagram illustrates methods for producing different versions of intelligent systems.
- **FIG. 6-1**: This diagram illustrates the concept of symmetric difference.
- **FIG. 6-2**: This diagram illustrates dimensions where data or other information might differ.
- **FIG. 6-3**: This diagram illustrates a specific method for determining the amount of valuable new information when comparing two datasets.
- **FIG. 6-4**: This diagram illustrates another method, based on Kaplan Information Theory, to evaluate the usefulness of information.
- **FIG. 6-5**: This diagram illustrates a general method for estimating information value.
- **FIG. 6-6**: This diagram illustrates a method for identifying useful information.
- **FIG. 6-7**: This diagram illustrates a method for identifying, acquiring, and simulating the effects of new information.
- **FIG. 6-8**: This diagram illustrates methods and heuristics that can accelerate the learning of AI systems.
- **FIG. 6-9**: This diagram illustrates a specific goal-related method that an intelligent entity might use to increase its intelligence.
- **FIG. 6-10**: This diagram illustrates an intelligent entity's method to find information that is maximally different from what the entity already possesses.
- **FIG. 6-11**: This diagram illustrates methods for validating the usefulness and safety of information.
- **FIG. 7-1**: This diagram illustrates a method that intelligent entities can use to achieve consensus on values.
- **FIG. 7-2**: This diagram illustrates a method that intelligent entities can use to achieve consensus on values.
- **FIG. 7-3**: This diagram illustrates methods that intelligent entities can use to identify, analyze, weight, and optionally combine ethical or other information to reach consensus.
- **FIG. 7-4**: This diagram illustrates a general method that intelligent entities can use to identify, elicit, and train on ethical information.
- **FIG. 7-5**: This diagram illustrates a method based on the principle of using converging evidence.
- **FIG. 7-6**: This diagram illustrates a method that intelligent entities can use to delegate voting authority.
- **FIG. 7-7**: This diagram illustrates a reputational process that intelligent entities can use to preserve information.
- **FIG. 7-8**: This diagram illustrates intelligent entities' methods to make sound ethical decisions.

- **FIG. 7-9**: This diagram illustrates a method that advanced AI can use to learn, test, improve, and monitor safety and regulation-related rules.
- **FIG. 7-10**: This diagram illustrates a method based on the Consequentialist Approach.
- **FIG. 7-11**: This diagram illustrates a method based on the Deontological Approach.
- **FIG. 7-12**: This diagram illustrates a method based on the Virtue Ethics Approach.
- **FIG. 7-13**: This diagram illustrates a method based on the Golden Mean Approach.
- **FIG. 7-14**: This diagram illustrates a method that trains an AI system to be human-aligned and compliant with regulations.
- **FIG. 7-15**: This diagram illustrates a method to align a customized foundation model or other AI system with specific expertise or group ethics.
- **FIG. 7-16**: This diagram illustrates a method to form an AGI or PI aligned with human ethics and values.
- **FIG. 8-1**: This diagram generally describes the current technology for online advertising systems.
- **FIG. 8-2**: This diagram illustrates a method for implementing a spot market.
- **FIG. 8-3**: This diagram illustrates a method for implementing the direct sale of attention.
- **FIG. 8-4**: This diagram illustrates a method for implementing an auction.
- **FIG. 8-5**: This diagram illustrates a system and methods for gathering expertise and cognitive work.
- **FIG. 8-6**: This diagram illustrates a method for improving online ad targeting.
- **FIG. 8-7**: This diagram illustrates a method for improving the effectiveness of the attention/expertise spot market.
- **FIG. 9-1**: This diagram illustrates the relationship between self-awareness and potential awareness.
- **FIG. 9-2**: This diagram illustrates the concept of multiple identities.
- **FIG. 9-3**: This diagram illustrates a general method for modeling awareness.
- **FIG. 9-4**: This diagram illustrates the minimum required components an intelligent entity must have to shift attention effectively.
- **FIG. 9-5**: This diagram illustrates the process for setting parameters for working memory.
- **FIG. 9-6**: This diagram illustrates a method for monitoring and updating awareness.
- **FIG. 9-7**: This diagram illustrates an attentional interrupt system.
- **FIG. 9-8**: This diagram illustrates general methods that an intelligent entity can use to gather input that can change the entity's sense of identity.
- **FIG. 9-9**: This diagram illustrates methods an intelligent entity can use to train the foundational model.
- **FIG. 9-10**: This diagram illustrates a "Turing Test" that can be used to determine when an intelligent entity has been trained sufficiently.
- **FIG. 9-11**: This diagram illustrates a method for arriving at a group identity.

- **FIG. 9-12**: This diagram illustrates another method for forming a group identity.
- **FIG. 9-13**: This diagram illustrates a method for resolving conflicts between identities.
- **FIG. 9-14**: This diagram illustrates a method that establishes, improves, and monitors behavioral protocols.
- **FIG. 9-15**: This diagram illustrates methods for simulation and consequence prediction.
- **FIG. 9-16**: This diagram illustrates a method for determining identity-based action.
- **FIG. 9-17**: This diagram illustrates a method for developing, refining, and evolving identities.
- **FIG. 9-18**: This diagram illustrates a general process for reasoning ethically and predicting consequences.
- **FIG. 9-19**: This diagram illustrates a method for resolving identity conflict using a process of hierarchical override.
- **FIG. 9-20**: This diagram illustrates an arbitration process method for resolving identity conflict.
- **FIG. 9-21**: This diagram illustrates a method for resolving identity conflict using negotiation and compromise.
- **FIG. 9-22**: This diagram illustrates a method for temporarily suspending and identifying (or identities) that may lead to destructive conflict.
- **FIG. X-1**: This diagram illustrates the evolution of PI.
- **FIG. X-2**: This diagram illustrates key dimensions of a PI system.
- **FIG. X-3**: This diagram illustrates an example of a method for automatically extending a network of AGIs.
- **FIG. X-4**: This diagram illustrates the architecture of PI.
- **FIG. X-5**: This diagram illustrates an exemplary PI architecture.
- **FIG. X-6**: This diagram illustrates the intelligence levels supported by the inventive methods disclosed in the white paper.

Importance of the White Paper

- The author states that the stakes for humanity are the highest ever in human history, and he warns that if humans do not act quickly to address the dangers of AI, then this century could be the last for humanity.
- The author believes that PI has the potential to either eliminate all forms of human poverty and material suffering or eliminate all forms of human life.
- The author's solution to this problem is to develop PI in a safe, ethical way that is aligned with human values.

- The author presents designs that ensure PI's safe and ethical development and embody the fastest and safest path to the development of PI.

White Paper #10 is significant because it proposes a novel and comprehensive approach to developing PI based on the principle of collective intelligence and incorporating a robust safety framework.

The author also provides an exemplary implementation of a PI system, based on the assumption that the implementor is a large technology company like Meta, Google, Nvidia, Microsoft, or OpenAI.

White Paper #10 is essential because it provides a roadmap for developing PI safely, ethically, and aligned with human values.

