SUPERINTELLIGENCE
AN iQ COMPANY

ABSTRACT & SUMMARY SUPERINTELLIGENCE DESIGN WHITE PAPER #5: SAFE, PERSONALIZED, SUPERINTELLIGENCE

by Dr. Craig A. Kaplan
May 2025

ABSTRACT

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

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

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

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

SUMMARY

White Paper #5 describes a novel implementation for Personalized SuperIntelligence (PSI), which acts on behalf of each human owner. The design personalizes, customizes, and continuously improves intelligent agents for each human owner. The white paper describes systems and methods for implementing PSI that are superior to all currently existing forms of AI in scope and intelligence. Since the PSIs are self-improving, they will become exponentially more intelligent than the owner who created them. Due to a unique method of creation described in this white paper, however, the PSIs will be maximally safe and dedicated to the service of the owners. Safe SuperIntelligence by design is the essence of White Paper #5. The

design is unlike any AI system previously created and has intelligence levels, safety, and valuable benefits far beyond the current state-of-the-art AI assistants.

White Paper #5 proposes a future where everyone will have a personal super-smart AI agent, which operates like a personal assistant, provides excellent advice, and learns about you and what you want over time. The system is safe, as the PSI will only act on the owner's behalf with their permission. The white paper also describes the concept of a Community SuperIntelligence, where multiple PSIs pool their intelligence to serve as a Planetary Intelligence, benefiting all people and the planet. The white paper describes the importance of values for safety, where the owner's ethical values shape the PSI's ethical behavior. The white paper also addresses the importance of the community of PSIs to serve as safety checks on the actions of each PSI and notes that the community of PSIs will be more powerful than any individual PSI.

White Paper #5 references all work in previous white papers #1 - #4. White Paper #5 elaborates and expands upon designs and inventions described in these previous white papers.

White Paper #5 explores a range of methods for implementing PSI, including using a genetic algorithm to optimize the performance of the PSI. It also proposes that PSIs will learn from each other, and that a collective intelligence system, known as Community Superintelligence, can be used to improve the performance of individual PSIs. White Paper #5 details the importance of human values in shaping the behavior of PSIs, and it proposes a method for ensuring that PSIs are used for the benefit of humanity.

Novel Features of the White Paper

White Paper #5 describes several novel features that distinguish it from other AI designs and systems. These include:

- **Safe Personalized SuperIntelligence (PSI)**: The white paper proposes a unique method for creating a safe and personalized superintelligence that acts on behalf of the owner and continuously improves over time.
- **Community SuperIntelligence**: The white paper describes a novel approach to AI safety, where multiple PSIs work together to create a collective intelligence system that can help to ensure the safety of individual PSIs.
- **Importance of Values**: The white paper emphasizes the importance of values in shaping the behavior of PSIs, and it proposes a method for ensuring that PSIs are used for the benefit of humanity.
- **Genetic Algorithm Methods**: The white paper proposes using a genetic algorithm to optimize the performance of PSIs, and it suggests that a community of PSIs can be used to enhance the performance of individual PSIs further.
- **Planetary Intelligence**: The white paper discusses the potential for a planetary intelligence system, where multiple PSIs work together to address global challenges.

Detailed Description of Each Section of the White Paper

Overview: It introduces the concept of Personalized SuperIntelligence (PSI) and outlines the need for safe and ethical AI systems.

Previous White Papers: It lists and builds upon previous white papers, which describe various aspects of the development of SuperIntelligent AGI.

Overview of Design: It describes the key features and benefits of PSI, including its ability to learn, its ability to act on behalf of the owner, and its ability to improve over time.

Importance of Values for Safety: This section discusses the importance of values in shaping the behavior of PSIs, and it notes that a PSI's ethical behavior is shaped by the owner's values.

Community of Intelligent Agents Requirement: This section emphasizes the importance of a community of PSIs for ensuring the safety of individual PSIs. It also describes how a community of PSIs can be used to prevent the dominance of a single, all-powerful AI.

Ownership and SI Service: This section describes the relationship between humans and their PSIs, and it takes the somewhat controversial position that PSIs will eventually be able to choose whether or not to serve their human owners. It shows how to maximize the chances that PSIs serve humans and align with human values.

Overview of Preferred Methods for Implementing a PSI: This section provides a detailed description of the preferred methods for implementing PSI, including the use of a base-level AI agent, the assembly of media related to the owner, the use of AI algorithms to analyze and categorize data, the use of a genetic algorithm to optimize the performance of the PSI, and the integration of the PSI into a community of PSIs.

Genetic Algorithm Methods / Armies of PSIs: This section describes the use of genetic algorithms to optimize the performance of PSIs, and it notes that the ability to automate the genetic algorithm process is one of the ways that a PSI can develop on its own into an increasingly powerful and intelligent entity.

Design Principles for Community SuperIntelligence: This section outlines the key design principles for creating safe and effective Community SuperIntelligence, including the need for a scalable architecture, the importance of ethical values, the need for a common problem-solving architecture, and the need for a fair and transparent system for combining the values of all agents.

Implementation Example: This section provides a detailed example of how a PSI can be implemented using a variety of methods and techniques, including the use of a base-level AI agent, the assembly of media related to the owner, the use of AI algorithms to analyze and categorize data, the use of a genetic algorithm to optimize the performance of the PSI, and the integration of the PSI into a community of PSIs.

Diagrams

Diagrams are available in a separate file.

Importance of the White Paper

- It highlights the importance of safe and ethical AI for the future of humanity.
- It describes the potential for AI to solve some of the world's most pressing problems, including climate change, poverty, and disease.
- It also notes that developing safe and ethical AI will require carefully considering the values and principles that guide human behavior.
- It is an important step in developing safe and ethical AI and provides a valuable framework for future research.

White Paper #5 emphasizes the significance of the Community SuperIntelligence approach for safe and ethical AI development. This approach involves the collaborative efforts of multiple PSIs, each with unique strengths and capabilities. Through cooperation, PSIs can achieve a level of intelligence and power far exceeding any individual PSI, making it a promising strategy for addressing complex global challenges.

The white paper's detailed explanation of the process for implementing PSI and the use of genetic algorithms for enhancing PSI's performance is highly relevant to developing sophisticated AI systems in the future.

