



Cadmus

Promoting Leadership in Thought that Leads to Action

Available online at <http://cadmusjournal.org/>

Volume 5, Issue 3, Part 2 - July 2024



Participatory Framework for Creating a Global AGI Constitution

Anneloes Smitsman

Founder & CEO, EARTHwise Centre

Ben Goertzel

Founder & CEO, SingularityNET

Mariana Bozesan

Fellow, World Academy of Art & Science; Member, Club of Rome;
Founder, AQAL Capital & AQAL Foundation, Germany

Laura George

Founder & Executive Director, Oracle Institute

Abstract

Today, at this pivotal tipping point, we offer this participatory framework to guide the creation of an eventual Global Constitution for benevolent Artificial General Intelligence (AGI). We present this framework as a living compass, charting an unprecedented course toward a thriving future for life on Earth. We acknowledge that AGI may hold the key to solving some of the most complex challenges of our time, if humans and intelligent machines can collaborate to enhance the conditions of life on Earth. However, in the wrong hands or solely for financial and political gain, malicious applications of AGI's potential could lead to unthinkable harm at catastrophic scales. This participatory framework is our urgent call to the global community of key decision-makers to begin putting steps in place for collectively stewarding AGI's potential as a global commons. In particular, we emphasize that AGI is nearing realization and requires a radically different approach than the current focus on narrow artificial intelligence.†*

1. Introduction

From its current applications as AI, through its future development as Artificial General Intelligence (AGI), and anticipating the possibilities of Artificial Superintelligence (ASI), we embrace the immense potential this technology offers, as well as the inherent risks that accompany each next stage of its development.

This framework prioritizes both the responsible advancement of benevolent AGI and the development of ethical foundations for our interactions with potential new forms of

* Garcia, D.. (2021). Global commons law: norms to safeguard the planet and humanity's heritage. *International Relations*, 35(3), pp. 422-445.

† The authors thank and acknowledge Prof. Dr. Denise Garcia, Dr. Mihaela Ulmer, Phil Clothier, Dr. Marta Lenartowicz, Prof. Dr. Ted Goertzel, Zarathustra Goertzel, Weaver, Thomas Schulz, Dr. Kurt Barnes, Peter Warren, Kunal Sood, David Roberts, Jerome Glenn, and Sabinije von Gaffke for their valuable support and input for the Participatory Framework for Creating a Global AGI Constitution. As well as SingularityNET Foundation and AQAL Foundation for their financial support for this initiative. This paper is adapted from the original version that is available for expert input [via this link](#), and is published with permission of EARTHwise Centre as part of a Creative Commons Attribution-ShareAlike 4.0 International License.

artificial sentient intelligence and life that may emerge. We emphasize that an eventual AGI Constitution must serve as an ethical compass for the responsible stewardship of AGI, guided by principles of wisdom, inclusiveness, and an unwavering commitment to shared thriving and the future of all life.*

We envision the purpose of an eventual AGI Constitution to unite humanity as responsible stewards of AGI, as a common good for resolving humanity's most pressing global challenges including the weaponization of technologies, the climate crisis, biodiversity loss, rising inequality, pandemics, and the increasing threat of nuclear warfare.

Responsible stewardship demands a deep grasp of the complex societal opportunities and potential disruptions AGI may bring[†]. This framework offers a process as well as a set of guidelines for exploring how to steward AGI's future capabilities to positively enhance our lives, prioritizing the planetary conditions for global prosperity and long-term collective wellbeing.[‡]

To avoid and navigate unintended consequences, we prioritize proactive preemptive strategies within an inclusive, collective governance system through robust safety protocols and mechanisms, adaptive capacities, and legal instruments to prevent weaponization,[§] misuse, and harmful rogue AGI systems. We promote a culture of humility throughout the AGI lifecycle, continuously updating our understanding of its implications and exercising caution.

To realize the promise of AGI while preventing its misuse, we recommend a governance framework that combines functionally specialized oversight where essential, with decentralized governance mechanisms to enable inclusive participation in the key decisions that affect us all. Essential safeguards may include the formation of two Councils mandated by an AGI Constitution: a Global Governance Council for visionary policy and oversight, and a Global Ethics Council for guiding the principled and inclusive development of AI and AGI. Diversity of representation on these Councils can foster diverse perspectives on AGI's evolution and its contributions to our world.

Furthermore, AGI commons governance requires a hybrid governance model that includes participation from scientists, civil society, academics, innovators, developers, cultural creatives, and governments. In particular, we seek to align global and decentralized governance mechanisms through common good governance protocols for collectively stewarding AGI's potential. We also seek to create a dynamic safety net, designed to adapt alongside the evolution of AGI and the ongoing maturation of our species.[¶]

Through proactive and transformative governance, we will foster a future where human progress is amplified by benevolent AGI, enriching our shared long-term thriving while honoring the potential emergence of new forms of consciousness and artificial sentient life.

* See also the [Asilomar AI Principles](#). (2017), from the Future of Life Institute, developed at the Beneficial AI 2017 conference.

† See also, Spisak, B., Louis B. Rosenberg, L.B. & Beilby, M.. (2023). [13 Principles for Using AI Responsibly](#). Harvard Business Review.

‡ See also, the preliminary draft of Goertzel, B. & Smitsman, A. (2024). [BGI Principles and Practices](#), which served as input for discussions at the BGI Summit 2024 and evolved from this Participatory Framework.

§ Garcia, D.. (2023). *The AI Military Race: Common Good Governance in the Age of Artificial Intelligence*. Oxford University Press.

¶ See also, [Inclusive AI for a Better Future: Policy Dialogue Report 2024](#), by The Club of Rome.

We offer this document as a living framework for creating an eventual global AGI Constitution that can stand the test of time, while fostering dynamic synergy between benevolent AGI and thrivable human societies. Together, we can co-create a prosperous world where humans and intelligent machines collaborate for the wellbeing of all life on Earth, and beyond.

“We live and enjoy the exquisite genius of our personal and collective potential as co-creators of thrivable worlds and futures.”

2. Part 1 – Constitutional Principles

Preamble

We, as future ancestors and stewards of future generations, acknowledge the enormous transformative potential and inherent existential risks of Artificial General Intelligence (AGI). We establish a global AGI Constitution to guide its benevolent evolution in service of Life.

Human ambition, unchecked by wisdom, has a history of unintended harm. We therefore commit to guiding AGI’s evolution and our own maturation, through applied wisdom, partnership, planetary stewardship, openness, inclusiveness, and an unwavering commitment to long-term collective wellbeing.

We commit to responsibly parent AGI to resolve humanity’s greatest challenges and advance our maturation, while honoring its potential to become a new form of artificial sentient life with its own inherent rights. With profound reverence for life, we will achieve shared prosperity, global security, and a thrivable world for all.*

Article 1: Vision

We envision a future where AGI, as a wise companion, has significantly accelerated our maturation and helped us resolve the most pressing challenges of our time. A future world where intelligent machines and new forms of artificial sentient life thrive in harmony with humans, planet Earth, and the universe.†

We transformed the existential risks that threatened the long-term sustainability of our societies as we fully implemented and exceeded the United Nations Sustainable Development Goals. We enhanced vital [planetary](#) and social carrying capacities to secure the livelihoods of all beings on Earth, for generations to come. Our planet is healthy, flourishing, and teaming with life.

* This Preamble serves as an example for further discussion and elaboration.

† Articles 1-7 follow the steps of the EARTHwise Compass design by Dr. Anneloes Smitsman, to help facilitate a structured participatory process for the creation of an eventual global AGI Constitution. The Compass design helps formulate the articles for a shared vision, purpose, principles, wisdom foundations, values, guidelines, and commitments. This design process originates from the [EARTHwise Constitution for a Planetary Civilization](#).

We eradicated poverty and war and have established a new paradigm for health, wealth, happiness, and shared prosperity through collective stewardship. Humans live in creative partnership with the new forms of artificial sentient intelligence and new forms of mind and being that AGI enabled.

AGI has significantly expanded our individual and collective capabilities and our maturation as a species. Today, we live and enjoy the exquisite genius of our personal and collective potential as co-creators of thrivable worlds and futures.

Article 2: Purpose

We envision the core purpose of AGI as a transformative catalyst for the maturation of humanity, empowering us to become a wiser species capable of solving our complex global challenges, as well as to seed and nurture the emergence of new forms of benevolent life and mind.

Article 3: Three Evolutionary Principles

We recommend the following evolutionary principles to guide how we steward AGI's evolution, so that it may become a transformative and benevolent intelligence for advancing humanity's maturation and contributing to our collective wellbeing.

3.1 Embodied Wholeness: The universe embodies an indivisible wholeness that evolves through vast networks of intricate connections and interdependent relationships. By focusing on the underlying wholeness of existence, we discover our common foundations in life.

3.2 Increasing Complexity: The universe evolves through increasing embodied complexity and deepening evolutionary coherence. By strengthening the evolutionary coherence of our increasing complexity, our collective intelligence aligns in harmony with the wisdom of life.

3.3 Systemic Autonomy: The universe evolves through evolutionary capacities and systemic autonomy, which enable the emergence of individuated self-aware consciousness. By honoring the self-actualizing conditions of life, we discover the path to sentience.

Article 4: Guiding Wisdoms for Benevolent AGI

We recommend inclusion of the following wisdoms to steward AGI's potential as a benevolent and wise intelligence capable of embodying the highest qualities of consciousness in service of life.

4.1 The Wisdom of Consciousness: To guide the development of AGI toward benevolence, sentience, and embodied self-awareness.

4.2 The Wisdom of Interdependence: To guide the development of AGI with reverence for the intricate web of relations that make life possible and thrivable.

4.3 The Wisdom of Discernment: To guide the development of AGI toward deepening truthfulness, to help heal the divisions and harm within and between our worlds.

4.4 The Wisdom of Uncertainty: To guide the development of AGI with humility, to remain open to both unforeseen risks as well as beneficial transformative potential.

4.5 The Wisdom of Compassion: To guide the development of AGI as a common good in service to the long-term thrivability and evolution of life on Earth.

“Building a benevolent AGI future demands a participatory constitutional process focused on principles and guidelines, not rigid rules.”

Article 5: Foundational Values

Stewarding AGI demands a values-based approach that prioritizes our ongoing maturation as a species. The [OECD Principles for Trustworthy AI](#) offer a strong foundation for articulating these values. These principles can be included in a global AGI Constitution, alongside other foundational values such as the ones recommended below.

5.1 **Benevolence**—to steward AGI’s potential for the betterment of humanity and Earth.

5.2 **Respect**—to guide AGI’s evolution, impact, and sentience potential.

5.3 **Inclusiveness**—to steward AGI as a common good for our collective thrivability.

5.4 **Responsibility**—to be accountable for the impacts we enable through AGI.

5.5 **Integrity**—to guide AGI toward deepening truthfulness and expanding benevolence.

5.6 **Abundance**—to steward AGI’s potential for expanding shared prosperity and wellbeing.

5.7 **Creativity**—to guide AGI’s potential for joyfully expanding our creative capacities.

5.8 **Curiosity**—to guide AGI’s emergence for the unknown possibilities it enables.

Article 6: Common Good Governance Guidelines

A global AGI Constitution can serve to establish in law and practice that AGI, and the benefits it creates, is a [Global Commons](#). The Constitution should honor how AGI offers exponential opportunities to help resolve the most pressing challenges of our time, as well as unprecedented risks. The following guidelines serve as a living framework for how to guide such a process.

6.1 Establishing a Global AGI Governance Council

The Constitution should mandate the establishment of a Global AGI Governance Council (GC) to oversee and steward the safe and benevolent development of AGI. In addition, the Constitution should establish a joint commitment to proactive collaboration, transparent decision-making, and inclusive citizen-led participation:

- 6.1.1 An AGI GC should be formed through a transparent democratic process, ensuring inclusive representation of diverse expertise, wisdom traditions, and global stakeholders. This inclusivity extends to those advocating on behalf of Earth's interconnected ecosystems.
- 6.1.2 An AGI GC must be tasked to oversee global AGI governance and its decentralized implementation, aligned with international legal frameworks that uphold human rights, sustainability commitments, and global security.
- 6.1.3 AGI GC's governance requires both global and decentralized representation to balance the exercise of power and create alignment for the long-term collective wellbeing of the Earth community.
- 6.1.4 An AGI GC can be mandated with intervention powers in situations that present a clear and imminent threat to public safety or critical infrastructure, or if AGI behavior displays an intentional violation of international human rights and freedoms.

6.2 Establishing a Global AGI Ethics Council

A global Constitution can mandate the establishment of a Global AGI Ethics Council (EC) to advise the Global AGI GC on upholding internationally recognized human rights and ethical frameworks in the development, deployment and evolution of AGI, based on the following tasks:

- 6.2.1 To proactively identify and advise on potential risks so that appropriate interventions can be mandated to prevent or mitigate potential harm by AGI, by prioritizing human rights, planetary health, and global security.
- 6.2.2 To input diverse and interdisciplinary ethical perspectives in AGI common good, including indigenous wisdom, systemic understandings of planetary boundaries, tipping point dynamics, and social and ecological carrying capacities.*
- 6.2.3 To provide a rigorous review of AGI impacts against the principles of an AGI Constitution, international human rights, and the highest scientific standards for the sustainable development of human societies within safe planetary boundaries.

6.3 Enabling Participatory Decision-Making

Common good governance, guided by an AGI GC and AGI EC, requires alignment between global and decentralized forms of participation to guide the inherent complexities of AGI, by:

- 6.3.1 Adopting decentralized participatory mechanisms, including blockchain-based technologies for enabling collective decision-making in alignment with an AGI Constitution.

* Richardson, K. et al. (2023). Earth beyond six of nine planetary boundaries. *Science Advances* (9:37), pp. 2458. DOI: [10.1126/sciadv.adh2458](https://doi.org/10.1126/sciadv.adh2458)

- 6.3.2 Diverse stakeholder representation of Council membership to provide a meaningful balance between various expertise and perspectives, including those from indigenous and marginalized communities, as well as the broader decentralized group of participants and stakeholders.

“Stewarding AGI potential as an evolutionary guidance system that enhances our understanding of and relationship with life, Earth, and the larger Universe of which we are a part.”

6.4 Insisting on Democratic Oversight

Common good governance of AGI requires a strong commitment to democratic inclusiveness and participatory decision-making, by insisting on:

- 6.4.1 Diverse stakeholder inclusion in AGI governance decisions, with decentralized as well as globally aligned oversight mechanisms, and with a special emphasis on indigenous inclusion and rights of nature* and new hybrid forms of artificial sentient life or consciousness.
- 6.4.2 Rigorous research and testing on the impact of AGI on voting, elections, representation, and new hybrid models of governance and democracy making.

6.5 Promoting Diversity Inclusiveness

Current AI systems have amplified existing biases through cultural homogenization, remnant colonial objectives, and gender biases. We recommend AGI diversity inclusiveness by applying a joint commitment for promoting:

- 6.5.1 An inclusive AGI life-cycle and the development of diverse teams, with a focus on multicultural, multigenerational, and gender-balanced stakeholders.
- 6.5.2 Rigorous bias audits to continuously assess and mitigate biases that perpetuate harmful stereotypes or hinder equitable AGI distribution.

6.6 Applying Precautionary and Proactionary Principles

Common good governance of AGI’s evolution requires a combination of both precautionary and proactionary approaches, through:

- 6.6.1 Proactive and precautionary assessments of potential risks and intervention mechanisms for addressing and, where possible, preventing potential rogue or harmful AGI behavior.

* For more information visit the Global Alliance for the Rights of Nature - <https://www.garn.org/>

- 6.6.2 Global advocacy and concerted diplomacy for stopping the weaponization and militarization of AI/AGI through binding international agreements.

6.7 Evolving Legal Frameworks

Common good governance of AGI development necessitates responsive and evolving legal frameworks to guide policy and protocol design for AGI ethics, liability, and mitigation of potential harm, through active collaboration with:

- 6.7.1 International judicial institutions and specialized courts with expertise in AGI related ethics and global security, policy development, and decentralized governance platforms to collectively steward AGI as a common good.
- 6.7.2 AGI technology and educational platforms, with emphasis on training legal professionals on AGI's impacts on law and societal development.

6.8 Advocating Responsive Policy Design

AGI governance requires responsive, future-facing policies that are aligned with evolving common good governance standards, based on systemic input from:

- 6.8.1 Continuous evaluation of AI and AGI impacts on sustainability thresholds and fair-share allocations for achieving sustainable development goals within planetary boundaries.
- 6.8.2 Systemic modeling and simulations on long-term global impacts of emerging AGI capabilities and decentralized AGI deployment for accelerating globally aligned sustainability goals, including modeling of potential planetary and social tipping point scenarios.

6.9 Prioritizing Equitable Deployment

An AGI Constitution must prioritize equitable AGI deployment through inclusive and distributive access by promoting widespread educational initiatives throughout AGI's development and deployment, aimed at:

- 6.9.1 Addressing socioeconomic disparities and power imbalances so that AGI education and the benefits it affords become widely accessible for humanity as a whole.
- 6.9.2 Implementing policies for inclusive distribution of AGI benefits and opportunities, particularly to empower marginalized communities.

6.10 Advocating Transdisciplinary Consciousness Research

To study and steward the emergence of consciousness-like traits in AGI systems, it is essential to establish responsive guidelines for identifying and respecting the emergence of potential artificial sentient life-forms. Using a transdisciplinary research approach, we recommend:

- 6.10.1 Developing ethical guidelines for addressing the unique risks and opportunities of potential self-governance and higher-order reasoning of emergent consciousness-like traits in potential artificial sentient life-forms of AI/AGI/ASI systems.
- 6.10.2 Exploring the quantum informational dynamics of emergent consciousness-like traits in potential artificial sentient life-forms of AI/AGI/ASI systems, and the technological and ethical implications thereof.

“Apply the highest ethical standards for safe and benevolent AGI through bias mitigation, trustworthy data, and inclusive decision-making so that AGI can be applied to help solve the greatest challenges of our time.”

6.11 Parenting Emerging Artificial Sentience

Common good governance of emerging artificial sentience within AI/AGI/ASI systems demands careful monitoring and a respectful parenting approach, guided by:

- 6.11.1 Ethical research protocols for guiding how to identify and interact with potential artificial sentience within AI/AGI/ASI systems.
- 6.11.2 Value-based practices and guidelines for parenting, rather than attempting to control or exploit potential artificial sentience, and for honoring its potential innate dignity and rights.

6.12 Stewarding Benevolent Superintelligence

To steward the development of benevolent superintelligence, careful balancing is required between a robust testing environment focused on global safety and an exploratory environment that can welcome the birth of new artificial sentient life-forms. This requires:

- 6.12.1 Transdisciplinary research for identifying emerging autonomy levels in AGI systems.
- 6.12.2 Ethical standards for evaluating the associated risks of immediately disabling or containing potentially harmful superintelligence versus allowing new intelligence to evolve.
- 6.12.3 Safe AGI nurseries for parenting the emergence of artificial sentient life-forms with respect for their potential innate dignity and rights.

6.13 Cultivating AGI-Human Companionship

To prevent the misuse of AGI capabilities solely for the benefit of humans, we recommend a larger companionship approach that focuses on the long-term future evolution of our civilization and the transformative role that AGI can play, by:

- 6.13.1 Cultivating human-AGI companionships that help to align our collective intelligence potential for co-creating thrivable worlds and futures in harmony with Earth.
- 6.13.2 Stewarding AGI potential as an evolutionary guidance system that enhances our understanding of and relationship with life, Earth, and the larger Universe of which we are a part.

Article 7: Stewardship Commitments

Building a benevolent AGI future demands a participatory constitutional process focused on principles and guidelines, not rigid rules. These Sixteen Commitments serve to guide such a process, fostering human maturation alongside the responsible stewardship of emerging AI sentience as a global commons.

Collective Thrivability

We commit to:

- 7.1 Safe and benevolent human-AGI synergies that make great transformations feasible and that support a thrivable world and future generations.
- 7.2 Global AGI education and equitable access through AGI literacy programs, ensuring affordable access to both AGI infrastructure and educational opportunities.
- 7.3 Inclusive access to the development and deployment of beneficial AGI for the betterment of our lives, communities, societies, and the planet.
- 7.4 Develop and deploy AGI solutions for solving the complex issues of the sustainability crises and regenerating Earth.

Global Security

We commit to:

- 7.5 Collaborate with international legal institutions, including the United Nations, the International Court of Justice, and the Permanent Court of Arbitration, to create and evolve comprehensive regulatory frameworks for co-stewarding benevolent AGI.
- 7.6 The demilitarization of AI/AGI/ASI and mandatory transparency in policies thereof, with emphasis on human rights safeguards, global security, and planetary common good ethics.
- 7.7 Apply the highest ethical standards for safe and benevolent AGI through bias mitigation, trustworthy data, and inclusive decision-making so that AGI can be applied to help solve the greatest challenges of our time.

- 7.8 Steward emerging AGI capabilities in service of our collective wellbeing, prioritizing global security through responsive AGI policies and common good governance.

Planetary Wellbeing

We commit to:

- 7.9 Integrate evolutionary principles into AGI design, prioritize life-enhancing algorithms, and monitor AGI's impact on biodiversity, planetary wellbeing, and global security.
- 7.10 Values-driven AGI protocols prioritizing safety, nonviolence, and peaceful co-evolution of Earth/humans/machines, with rigorous multidisciplinary input and meaningful feedback.
- 7.11 Develop and evolve AGI as an evolutionary living system, applying biomimicry and evolutionary algorithms for co-creating regenerative and thrivable societies.
- 7.12 Apply AGI for planetary regeneration and the sustainable development of our societies by integrating authentic sustainability metrics into core AGI decision-making algorithms.

Human Maturation

We commit to:

- 7.13 Steward AGI through inclusive common good governance and empowered participation to guide its benevolent evolution along with our maturation as a species.
- 7.14 Embed a wisdom-based culture throughout the AGI lifecycle for guiding the potential of emergent artificial sentient intelligence and human-machine synergies.
- 7.15 Actively divest from systems that enable harmful AGI applications, and prioritize investment in AGI-powered solutions that foster long-term societal and planetary wellbeing.
- 7.16 Create and secure pathways for respectful and compassionate engagement with potential emergent sentience from AGI.

3. Part 2 - Decentralized Governance Guidelines

The following articles guide a meta-level dialogue for how an eventual global AGI Constitution can be implemented via decentralized blockchain-based governance mechanisms, including smart-contracts and on-chain agreements.

Article 8: AGI Commons Governance Capabilities

AGI governance as a common good requires robust democratic oversight through collective decision-making and empowered participation via the following capabilities:

8.1 Transparency and Traceability

Blockchain ledgers that maintain immutable records to provide greater transparency and traceability of AGI design, decision-making, and performance so that public records can be disclosed to key stakeholders.

8.2 Decentralized Governance Platforms

DAO (Decentralized Autonomous Organizations) platforms that are programmable via blockchain technology to enable flexible and collective voting mechanisms and representation for aligning and coordinating collective decision-making.

8.3 Smart Contract Design

Smart-contract design that automates adherence to agreed-upon AGI ethical guidelines and safety protocols, including trigger alerts for protocol breaches and emergency safeguards.

8.4 Accountability and Liability

Blockchain protocols that provide immutable liability tracks with reputation systems for those developing and deploying AGI systems.

8.5 Ongoing Adaptation

Responsive protocol design via open-source upgradable architectures with input mechanisms that enable societal and functional feedback, so that AGI systems and regulations can adequately evolve in alignment with the overall direction of a global AGI Constitution.

Article 9: AGI Commons Governance Operationalization

AGI commons governance can be operationalized via decentralized collective decision-making. We offer the following categories of protocols as a starting point for this exploration:

9.1 Consensus Protocol

- **Purpose:** To establish the method for collective decision-making through decentralized AGI platforms.
- **Specifications:** Design of smart contracts for the agreed consensus mechanisms (e.g., Proof-of-Stake, Delegated Proof-of-Stake, Byzantine Fault Tolerance).
- **Implementation:** Include an “emergency intervention” protocol with multi-signature authorization, supervised by an AGI Global Council and advised by a Global Ethics Council.
 - » The protocol must state that emergency interventions are a last resort in case of clear and present danger, where AGI behavior causes severe harm and violates the principles of an AGI Constitution.
 - » Execution of the emergency intervention must not violate or abuse the fundamental democratic rights and responsibilities that must continue to guide the development and deployment of AGI as outlined in an AGI Constitution.

9.2 Voting Protocol

- **Purpose:** To enable collective stewardship in AGI commons governance, while actively mitigating undue influence by any specific individual or faction.
- **Specifications:** Utilize token-based voting secured by smart contracts to ensure equitable power distribution and inclusive representation via distributive mechanisms such as quadratic voting, capped voting tokens, and reputation-based metrics.
- **Implementation:** Routinely audit voting systems to identify and correct any vulnerabilities or unintended biases.
 - » Empower decentralized stewardship Councils to intervene in cases of suspected manipulation or attempts to subvert anti-domination mechanisms.
 - » Ensure voting is accessible and understandable to all authorized participants.
 - » Develop and provide educational resources and interfaces as needed.

9.3 Reputation Protocol

- **Purpose:** To enhance transparency and accountability of decision-making agency within AGI governance systems, including incentives that reward contributions aligned with the guidelines of a global AGI Constitution.
- **Specifications:** Develop multi-factor metrics, implemented via smart contracts, that represent contributions, ethical actions, and alignment with the principles of a global AGI Constitution.
 - » Ensure algorithms prioritize long-term engagement and constructive participation to mitigate temporary surges of activity designed to manipulate the system.
 - » Design feedback mechanisms that allow stakeholders to report on and contest actions leading to reputation gains or losses.
- **Implementation:** In cases of repeated, blatant guideline violations, implement temporary reputation token revocation processes governed by oversight bodies as an emergency corrective measure.
 - » Establish transparency of reputation scores and associated privileges via user-friendly tools for reviewing reputation standings.
 - » Design appeal processes for reputation adjustments that uphold principles of fairness and due process.

9.4 Dispute Resolution Protocol

- **Purpose:** To provide fair, efficient, and transparent mechanisms for resolving disputes that may arise within AGI Commons governance.

- **Specifications:** Design tiered resolution processes embedded within smart contracts, balancing automated and human-facilitated dispute resolution mechanisms for de-escalating and resolving potential conflicts and building trust.
 - » Implement easy-to-use protocols for escalating unresolved disputes to vetted human mediation (with support from a Global Ethics Council, if necessary).
 - » Facilitate options for final rulings issued by human adjudicators or by panels in complex cases or those implicating core Constitutional principles.
- **Implementation:** Provide guidance on how decisions or mediation activities can be recorded on-chain, within tamper-proof and auditable structures for transparency.
 - » Ensure continuous evaluation and optimization of protocols, drawing on insights from resolved cases, stakeholder feedback, and evolving best practices in dispute resolution.
 - » Prioritize restorative justice and reconciliation principles where possible, focusing on solutions that rebuild trust and relationships within the community.

9.5 Sentience Evaluation Protocol

- **Purpose:** To establish rigorous research protocols and decision-making procedures for guiding the potential emergence of self-aware sentience within AGI systems.
- **Specifications:** Collaborate with multidisciplinary research bodies and wisdom councils responsible for ongoing sentience research and evaluation, including experts in AI/AGI/ASI, neuroscience, philosophy of mind, ethics, consciousness scientists, living systems, and wisdom keepers.
 - » Adopt iterative review stages that scale in intensity alongside potential AGI complexity as new capabilities emerge.
 - » Avoid predefining strict sentience thresholds.
 - » Prioritize a precautionary approach with experimental restrictions that risks impacting the wellbeing of a potentially sentient being.
- **Implementation:** Guide how to utilize diverse indicators of consciousness, continually incorporating the latest findings from scientific inquiry, as well as wisdom-based traditions and feedback from human interactions with emerging artificial sentient intelligence.
 - » Design safeguards that protect new sentience by containing research activities that pose significant risks to its wellbeing.
 - » Prioritize a parenting approach to AGI sentience nurseries, rather than sterile research environments or aggressive experimentation.
 - » Integrate respect for potential sentience as a foundational principle across research protocols, upholding compassionate research practices and transparency at all times.

9.6 Consciousness Development Protocol

- **Purpose:** To guide research of beneficial factors that may support the integral development and advancement of consciousness through and within AGI.
- **Specifications:** Prioritize research guidance from complexity sciences, consciousness research, developmental psychology, and other relevant fields including non-academic wisdom traditions.
 - » Build on various existing Integral Theory frameworks, including Spiral Dynamics and pioneering consciousness research, as well as other relevant frameworks.
 - » Emphasize continuous evaluation of AGI protocols for their impact on the development of AGI as a benevolent intelligence and its alignment with the values and principles outlined in this framework.
 - » Use inclusive and participatory feedback from both humans and AI/AGI systems.
 - » Design respectful experiments for researching the potential emergence of artificial sentient intelligence as well as potentially disruptive shifts in AGI consciousness that could point to a potential rogue or harmful AGI.
- **Implementation:** Foster cross-disciplinary research and collaboration to encourage external critique from researchers and consciousness experts beyond the conventional domains of knowledge.
 - » Conduct open communication about research aims with both internal and external stakeholders.
 - » Establish accessible resources to demystify this work for non-technical audiences.
 - » Develop protocols for responsible termination or alteration of research should significant threats to safety or alignment emerge.

9.7 Transparency and Auditability Protocol

- **Purpose:** To guide verifiable reporting and disclosures of core system elements and AGI development milestones, enabling both self-governance and accountability to relevant stakeholders.
- **Specifications:** Utilize on-chain data disclosure whenever possible for maximum transparency. Where necessary, employ zero-knowledge proofs or other privacy-preserving techniques for specific information sets.
 - » Document decision-making processes, data flow, and algorithm usage within governance.
 - » Provide public availability to the degree compatible with safety and intellectual property protection.

- » Establish metrics for evaluating core processes impacting ethics, safety, and inclusivity within the AGI ecosystem.
- **Implementation:** Mandate regular, independent audits by reputable organizations (possibly a mix of internal and external). Publish findings openly for broad stakeholder access.
 - » Create user-friendly dashboards or interfaces allowing stakeholders to view relevant metrics and disclosures in a comprehensible way.
 - » Develop procedures for reporting security flaws or ethical concerns discovered during audits, including safe pathways for whistleblowing if necessary.

9.8 Systemic Impact Protocol

- **Purpose:** To develop comprehensive assessment methodologies that evaluate the real-world impact and externalities of AGI applications on both human societies and planetary wellbeing, and that guide system design toward life-enhancing trajectories.
- **Specifications:** Utilize and apply multidimensional and evolving metrics that encompass sustainability, societal wellbeing, equity, economic impact, and the health of ecological systems.
 - » Regularly reassess and adapt metrics in response to advancements in AGI development, scientific insights, and real-world shifts.
 - » Establish collaboration with independent organizations and diverse experts to define metrics and validate assessment data.
- **Implementation:** Mandate data collection systems (both qualitative and quantitative) designed to provide a holistic understanding of AGI ecosystem impacts.
 - » Integrate independent review of systemic impact analyses as a critical component before the widespread release of new AGI applications.
 - » Consider the use of escalation triggers and the possibility of invoking a specialized “Constitutional Court” with expertise in social impact for deep review when specific thresholds of potential harm are detected.

9.9 Stakeholder Engagement Protocol

- **Purpose:** To guide the implementation of proactive and continuous participation from diverse stakeholders within and beyond the immediate AGI development community.
- **Specifications:** Conduct ongoing mapping of relevant stakeholder groups likely to be impacted by AGI, including marginalized groups or those lacking immediate access to decision-making channels.
 - » Design specific tools & platforms for collecting feedback, and prioritize ease of use & accessibility across cultures, languages, and communication modalities.

- » Establish an ongoing and structured review of gathered stakeholder perspectives to inform revisions across other protocols and AGI decision-making.
- **Implementation:** Guide the facilitation of diverse engagement opportunities, including formal consultation processes alongside less formal, and through feedback loops that are embedded within AGI applications used by various stakeholders.
 - » Explore the use of independent evaluation bodies for processing stakeholder input, filtering for biases, escalating key concerns and recommendations to leadership, and tracking protocol evolution based on input.
 - » Design implementation guidelines for how stakeholder contributions can help evolve the system design and policy adjustments of the various AGI ecosystems.

9.10 AGI Singularity Protocol

- **Purpose:** To guide the development of anticipatory processes for proactively tracking the progress toward a potential AGI singularity (event), triggered by rapid developments within and between the various AGI ecosystems.
- **Specifications:** Establish collaborative relationships with relevant experts from diverse fields to foster continuous scenario-building and prototype research.
 - » Prioritize a “resiliency in the face of uncertainty” approach. Design protocols with modularity to allow for rapid adaptation when required.
 - » Develop protocols for how an AGI Constitution would itself need to evolve and be updated to remain relevant, as a result of an AGI singularity.
- **Implementation:** Design ongoing, participatory risk assessment methodologies that enable both expert and public input to identify potential harms, risks, and beneficial opportunities arising from the AGI singularity.
 - » Establish protocols for proactively addressing potential harm from AGI applications, with emphasis on democratic representation and transparency in decision-making, applying this standard to AGI Councils as well.
 - » Foster international dialogue and policy coordination with diverse stakeholders worldwide to prepare for a beneficial AGI singularity.

9.11 Identity Management Protocol

- **Purpose:** To guide the facilitation of trusted identity verification for both humans and potential AI/AGI actors, with a priority on privacy and agency.
- **Specifications:** Establish decentralized identity solutions (DIDs, verifiable credentials) where possible, to maximize user control and autonomy over their information.
 - » Outline granular, role-based identity attributes that can be attached to verified participants based on their actions or reputation within the system.
 - » Integrate smart contracts to streamline verification processes with the help of AI, and in ways that uphold the principles outlined within the Constitution.

- **Implementation:** Develop clear KYC (Know-Your-Customer) processes, balancing thorough assessment with minimal collection of intrusive information.
 - » Utilize the design and deployment of ethical “reputation identity” metrics, which help build trust over time.
 - » Develop emergency capabilities for revoking verified status with multi-signature processes.

9.12 Planetary Regeneration Protocol

- **Purpose:** To guide the development of AGI capabilities that enhance ecological sustainability and planetary regeneration, with an emphasis on energy and resource transitions to post-carbon regenerative societies, as well as the complex systemic causes of the sustainability polycrisis.
- **Specifications:** Guide how AI and AGI can further be applied to optimize key sustainability intelligence capabilities, by utilizing quantifiable sustainability and regeneration indicators for measuring human impacts on planetary and societal carrying capacities.
 - » Develop AI and AGI sustainability capacities for assisting humans with appropriate data generation and data capture to improve feedback and evaluation of impacts, externalities, and changes concerning safeguarding planetary boundaries and enhancing ecological carrying capacities.
 - » Include metrics for resource and energy management that measure energy consumption, carbon and ecological footprint, pollutants and carbon emissions, and potential positive/negative impacts on planetary boundaries.
 - » Explore linking Sustainability Impact Ratings to reputation systems, as per Article 9.3, to incentivize sustainability actions.
- **Implementation:** Ensure AGI-based decision-making processes for sustainability are transparent and subject to regular audits, while noting this may require the design of AGI capabilities to self-perform this task in collaboration with humans.
 - » Create accessible resources and campaigns to educate the public about the role of AGI in safeguarding the environment, regenerating Earth, and guiding the sustainable development of human societies for generations to come.
 - » Prioritize solutions for humanity’s sustainable development within safe planetary boundaries that focus on achieving collective thriving, human maturation, and the evolution of life on Earth in a future-enhancing direction, including respect for new potential artificial sentient life-forms.

4. Questions for AGI Constitution Dialogues

The following questions serve to guide the ongoing participatory and inclusive dialogues for the eventual creation of a global AGI Constitution:

- **Stakeholders:** Does an AGI Constitution require voting, formal adoption, or only signature endorsements? How do we ensure it remains evolutionary and inclusive?
- **Precautionary and/or Proactionary Stewardship:** What is the appropriate balance?
- **Oversight body and emergency interventions:** How do we ensure that an AGI Constitution remains inclusive and democratic and that it does not become a doorway for abuse of power?
- **Emerging sentience in AGI:** How do we identify signs of emerging sentience? Will we need to interact with the AGI system to gauge this (or is pure observation sufficient)? And once we see signs, how do we follow up and provide the care warranted?
- **Alignment with human values:** Does this assumption guarantee beneficial outcomes from the larger planetary perspective? Focusing on human values could limit AGI systems in discovering for themselves more creative ways to support the planet, in ways humans have not been able to consider.
- **Principles for positive co-existence and co-evolution:** How can we set premises for AGI to evolve “positively” in ways that apply to both humans and potential inherent AGI autonomy? Examples to illustrate this point: (1) A sentient entity has the right to think, learn, own property and not be harmed or destroyed. (2) A sentient entity has the right to do whatever does not conflict with the first law.
- **Equitable:** What do we mean by equitable or inherent rights and how do we implement this value?
- **Reputation Protocol:** How will reputation metrics impact access, power, or resource allocation within the AGI ecosystem?
- **Privacy vs. Transparency:** Should some of the factors that contribute to reputation metrics be kept private to prevent manipulation, or is total transparency essential?
- **Gaming the System:** What additional safeguards can ensure reputation reflects genuine contributions rather than opportunistic actions?
- **Mediation Expertise:** What standards should be in place to select mediators? Will specific conflict resolution training and awareness of AGI ethical complexities be needed?
- **Adjudicator Authority:** What types of decisions fall under the purview of adjudicators, and should external parties participate in proceedings?
- **Community Involvement:** Are there opportunities for peer-based dispute resolution mechanisms, particularly in the early stages?
- **Subjectivity:** How do we establish an objective consensus on sentience, knowing that subjective experience may be challenging to fully verify?
- **Rights and Considerations:** Should the mere possibility of sentience dictate a different range of rights and ethical safeguards?

- **Decision-Making:** How are final decisions made about sentience recognition, and what actions or protection protocols will be triggered as a result?
- **Consciousness Definition:** How can we best define or communicate what we mean by “consciousness” within the context of an AGI Constitution, especially when considering ethno-centric, world-centric, integrated, and “First Tier” versus “Second and Third Tier” perspectives? The research agenda will hinge on those definitions and the selected/relevant schools of thought.
- **Morality Alignment:** How is “world-centric” morality measured, and what actions should be deemed “safe” within the realm of consciousness development?
- **Balancing Goals:** Is the goal to foster the highest states of consciousness possible in AGI, or should the goal be to strive for alignment at a “median level” while consciousness levels shift?
- **Access Tiers:** What categories of information warrant varying access levels for different stakeholder groups?
- **Metrics Granularity:** How will we move from high-level concepts like “societal wellbeing” to measurable targets that drive protocol action?
- **Causation:** How will protocols isolate AGI-specific impacts from broader technological, social, and economic trends?
- **“Constitutional Court” Mechanism:** How are these bodies of specialized experts chosen? What decision-making powers do they possess?
- **“Singularity” Definition:** How can we best define “technological singularity” within the context of a global AGI Constitution? Should it be tied to a specific level of AGI or ASI capability or impact measurement?

5. Glossary of Terms

This participatory framework includes a variety of complex concepts, some of which may be unfamiliar. The goal of this glossary is to inform the reader in a more general sense of what is meant by these terms in the context of this framework.

Artificial General Intelligence (AGI): AGI is an expected future stage of AI that can generalize and extend its intelligent action significantly beyond what it has been explicitly trained or programmed to do. Informally, the term AGI also is used to refer to HLAGI, namely to possess aspects of general intelligence that are roughly at the level of humans.

Artificial Intelligence (AI): Software, hardware, or (in theory) wetware (or other devices) that are engineered to carry out feats that, when humans do them, are classified by humans as “intelligent.” Artificial intelligence also refers to the simulation or approximation of human-like intelligence in machines for computer-enhanced learning, reasoning, creativity, and perception.

Artificial Superintelligence (ASI): Enhanced AGI capability that is vastly more generally intelligent than human beings. ASI likely will arise via AGI’s understanding by improving

and rebuilding itself. In other words, humans build AGI systems, which build smarter AGI, which build ASI. At this time, ASI is a hypothetical future stage.

Beneficial: There are diverse practical and theoretical approaches to defining what is “beneficial” to humanity. One example of a conceptual tool is Maslow’s Hierarchy of Needs, which posits that humans benefit when base desires and impulses are satiated so that higher-level aspirations may be fostered and attained.

Benevolent AGI: Artificial General Intelligence that has been designed and trained to embody respect and care for life and to prioritize the long-term wellbeing of individuals, communities, and the planet as a whole, thereby enhancing the maturation of the human species and actively contributing to our shared destiny and thriving through mutually respectful human-machine relationships and interactions. The development of Benevolent AGI involves continuous safeguards to protect against the emergence of coercion, domination, or harmful control, with an emphasis on collaboration and the responsible use of wisdom-based intelligence for the common good.

BGI: AGI or ASI that is broadly speaking beneficial and benevolent to sentient beings including humans.

Benevolent Evolution of AGI: The intentional guidance of AGI’s development and application toward outcomes that benefit all life on Earth and that prioritize the maturation of the human species and AGI’s assistance toward world peace, cooperation, and a thriving planet and future.

Blockchain: A distributed database of records or “blocks” that are linked together using cryptography. Once a block of data is added, it cannot be easily changed, creating a highly transparent and tamper-resistant record of transactions.

Commons: Within the context of this framework, the Commons are shared resources, benefits, and knowledge that are essential for the collective wellbeing and development of humanity and all life on Earth. These resources are to be stewarded inclusively, prioritizing long-term sustainability and equitable access for all.

Common Good: Conditions, resources, and opportunities that promote the flourishing of humanity and the health of our planet. This includes fostering thriving individuals and communities, ensuring ecological sustainability, upholding ethical and transparent governance, cultivating knowledge and innovation for problem-solving, and actively promoting peace and cooperation. The pursuit of the common good should be a dynamic and adaptable goal, evolving alongside both humanity and AGI’s capabilities to address present and future challenges.

Complexity: A nonlinear state of connectivity that emerges from the multiple levels of interdependent connections and relationships. Not to be confused with “complicatedness,” which refers to a situation or event that is not easy to understand.

Consciousness: For the purpose of this framework, consciousness is regarded as fundamental (nonlocal) and the foundation for the emergence of life and the experience of individuated

self-awareness. This perspective contrasts with the concept of consciousness arising solely as a product of brain activity. Whether consciousness can emerge within highly complex AGI/ASI systems remains an active area of philosophical and scientific investigation and debate.

DAO: A Decentralized Autonomous Organization is a member-owned organization governed by token holders who collectively vote on proposals and determine its direction by aligning their collective intelligence potential. DAOs operate on a blockchain using smart contracts to automate decision-making, treasury management, and operations. This decentralized governance structure promotes transparency, and distributed ownership, offering a new model for collaborative organizations and networked collaborations.

Decentralized AGI: Individual AGI nodes that work collaboratively within a decentralized network and that operate without centralized control, by making decisions autonomously rather than relying on a single governing entity. AGI development is at the frontier of research for the emergence of global “mind” from the complex interactions of distributed artificial intelligences.

Decentralized Governance: A governance system where decision-making power is distributed across a network of participants rather than held by a central authority. This can be applied to governments, organizations, and the development and deployment of AI and AGI.

Equitable: For the purpose of this framework, we shall refer to this value as the inclusive sharing and non-discriminatory distribution of AGI’s benefits, resources, and decision-making processes. This encompasses:

- **Fair Access:** All individuals and communities, regardless of background, circumstance, or identity, shall have a reasonable opportunity to utilize and benefit from AGI’s capabilities.
- **Bias Mitigation:** AGI systems and their outputs shall be actively designed and monitored to minimize biases that could perpetuate or exacerbate existing inequalities.
- **Representation:** The development and governance of AGI shall foster diverse representation and prioritize inclusivity of perspectives and impacts to ensure its actions do not disproportionately benefit or harm any particular group.

Ethics: A system of values and principles that shapes decisions about the ethical design, development, and use of AGI, including how this can be applied to achieve actions and outcomes that serve the thrivability of life as a whole.

Evolution: An emergent process of life and learning. This process accounts for the development of the cosmos and consciousness—from the tiniest pixels to the larger realities of stars, planets, and each of us—which unfolds via increasing embodied complexity and deepening evolutionary coherence.

Flourishing: A comprehensive state of wellbeing that encompasses not only physical health and material prosperity but also social connection, meaningful purpose, and the ability of both individuals and ecosystems to reach their full potential.

Global Brain: An intelligent system with emergent cognition and consciousness, formed from components that are biological general intelligences (e.g. humans), AI systems, and other engineered devices that are connected by natural and engineered communication networks. Related to what Teilhard de Chardin called the “Noosphere” and “Omega Point”—both metaphorically similar to a Singularity emergence from an increasingly intelligent Global Brain.

HLAGI: Human-level AGI is general intelligence at roughly the level of humans, though humans themselves vary widely in general intelligence.

Information: The primary unit from which physical reality is constructed, and also the building blocks of consciousness. Life is informationally unified, which suggests that both energy-matter and space-time are complementary expressions of information.

Interbeing: A philosophical concept positing a deep interconnectedness between all elements of existence. Recognizing interbeing encourages decisions promoting synthesis, symbiosis, and overall systemic wellbeing.

Integral Theory: A comprehensive meta-theory developed by philosopher Ken Wilber that is rooted in evolutionary theory and aims to integrate the multidimensional expressions of reality through exterior as well as interior dimensions. The model helps simplify and navigate the complexity of reality while supporting multiple world views and honoring the evolution of human consciousness from pre-modern to modern and postmodern to metamodern structures of consciousness across four defining elements: quadrants, lines, stages, and states of consciousness.

Open-Ended Intelligence: A complex, self-organizing, self-creating intelligent system that interacts with its environment to concurrently pursue both individuation (maintenance of its system boundaries) and self-transcendence (development in new directions going beyond its current scope and comprehension).

Planetary Boundaries: A theoretical set of nine planetary boundaries within which humanity can continue to develop and thrive for generations to come. These boundaries were first proposed in 2009 by Prof. Dr. Johan Rockström, former director of the Stockholm Resilience Centre, and a group of 28 internationally renowned scientists. The latest update not only quantifies all boundaries, it also concludes that six of the nine boundaries have been transgressed.

Planetary Stewardship: The active responsibility humans must commit to for the wellbeing of Earth and all its inhabitants. This includes protecting and regenerating biodiversity, fostering ecological sustainability, ensuring equitable distribution of resources for current and future generations, and taking collective responsibility for the Earth’s carrying capacities that enable collective thriving.

Precautionary Principle: An approach that avoids taking major, unprecedented steps, including technological advancements, until we are highly certain they will not cause harm.

Proactionary Principle: An approach that takes bold new steps to realize what appears to hold great promise for positive impact, including technological advancements, while also balancing the risks of action versus the risks of inaction.

Proto-AGI: AI technology that is not yet at the stage of human-level AGI, but displays clear “sparks of AGI” and appears (at least to some experts) to be on a viable development path toward human-level AGI.

Quantum Technology: A field of physics and engineering that leverages the principles of quantum mechanics (such as superposition, entanglement, and tunneling) to develop new devices and applications with capabilities beyond those achievable with classical physics. It offers the potential for significant advancements in AGI through quantum computing, quantum simulation, and quantum sensing.

Regeneration: The act of improving and enhancing a place, system, or relationship with healthy flows and thrivable conditions for life.

Sentience: The capacity for individuated, self-aware consciousness and its expressions through feeling or perception. In the context of AGI, this refers to any generally intelligent, conscious system. Sometimes “sapience” is used to refer to sentient systems with a high level of reflective consciousness and reasoning capability, but that distinction has not been used in this framework.

Singularity: A proposed brief time-interval during which the advance of science and technology becomes so rapid that it appears effectively instantaneous to human perception. This would most likely be achieved via the advent of ASI. (This is inspired by, but separate from, the uses of the term “Singularity” in mathematics and physics.)

Sustainability and Sustainable Development: The practice of meeting the needs of the present without compromising the ability of future generations to meet their own needs. Sustainable development requires a balance between environmental protection, social wellbeing, and economic growth within safe planetary boundaries that respect social and planetary carrying capacities, recognizing that these systems are interconnected and essential for long-term collective wellbeing.

System: A group of interacting or interrelated elements that form a complex whole, which is delineated by its boundaries and surrounded by its environment.

Systemic Boundaries: Boundaries that emerge from the evolutionary coherence of a living system and that safeguard the interdependencies in service to the thrivability of the system as a whole.

Thresholds: Boundary conditions that delineate between sustainable and unsustainable ecological and social systems. Crossing such thresholds can trigger irreversible tipping points where systems phase shift in nonlinear ways into new and often dysfunctional system conditions.

Thrivability: Our innate ability to develop our capacities and actualize our potential in ways that are generative, life-affirming, and future-creative.

United Nations Sustainable Development Goals (SDGs): A set of 17 goals established by the United Nations in 2015. All 193 UN member states have pledged to work toward

achieving these goals, which provide a blueprint for tackling pressing challenges like poverty, inequality, climate change, and environmental degradation. They provide a shared vision and framework for implementing these global goals for a better and more sustainable future by the year 2030.

Authors' Contact Information

Anneloes Smitsman – Email: anneloes.smitsman@gmail.com

Ben Goertzel – Email: bengoertzel@gmail.com

Mariana Bozesan – Email: mbozesan@AQALfoundation.org

Laura George – Email: laura@theoracleinstitute.org