



## Zero-Sum Power Politics vs. Synergetic Politics for Human Security

**Jerome C. Glenn**

CEO, The Millennium Project  
Fellow, World Academy of Art and Science

### Abstract

*Zero-sum power politics is a fundamental threat to human security. Synergetic relations among nations are proposed to improve human security. Among existential threats to humanity, the most immediate and little understood is the development of Artificial General Intelligence (AGI) before agreements are in place for its management. Examples of potential beneficial initial conditions, rules, and guardrails for AGI and governance modes are proposed.*

\*\*\*

*If the world continues to play zero-sum power geopolitics it seems continuing wars in one form or another is inevitable for our future.*

### 1. Synergetic Thinking

Synergy is a concept made popular by futurist R. Buckminster Fuller (Bucky) Fuller. He would say “you put a wheel in a box and you don’t get much, but put it under the box and you get a wheelbarrow and you can get plenty of work done.” Hence, it is not the parts that may create synergy, but their relationship. It is not that the wheel cooperates with the box, instead the synergetic relationship creates a new entity with properties not easily predicted by the parts.

What synergetic relationships between Taiwan, China, and the US are possible? What synergetic relationships could be created between India and the United States or China? Could a recovering Sri Lanka work with India to create international synergy?

The Millennium Project, a global participatory think tank, has just created the South Asia Foresight Network (SAFN) to conduct collaborative futures research for the region and explore synergies among these nations. While cross-border, maritime security, trade, economic and climate challenges will be at the forefront for many policymakers, it is important to find innovative synergetic solutions to move away from zero-sum mentality.

---

*“We need a  
United Nations  
Convention on  
Artificial General  
Intelligence.”*

---

What are the future potential synergies among the nations in the region? A synergetic matrix could be created, as shown below.

*Table 1: Synergetic Matrix*

	India	Pakistan	Nepal	Srilanka	Afghanistan	Bangladesh	Bhutan	Maldives
India	xxx	1	2					
Pakistan		xxx						
Nepal			xxx					
Srilanka				xxx				
Afghanistan					xxx			
Bangladesh						xxx		
Bhutan							xxx	
Maldives								xxx

To fill out cell 1, answer: what are possible synergetic relations of India with Pakistan; to fill out cell 2, answer: what are possible synergetic relations of India with Nepal; etc.

---

*“We have to create a global agreement on how to govern AGI BEFORE it is created. This could be the most difficult management challenge we face today.”*

---

University Schools of International Affairs should teach potentials for synergetic relations and analysis as well as current zero-sum power politics and competitive advantage. For example, we need a United Nations Convention on Artificial General Intelligence. The synergetic relationship between the United States and the People’s Republic of China could be to create an Apollo-like climate change goal and a NASA-like R&D Program to achieve it that others could join. This new organization would give hope to the world and focus research and policy on addressing one of the greatest threats to humanity.

Another synergetic relationship, instead of the current zero-sum politics, could be that the United States and China jointly introduce a UN General Assembly resolution to create the UN Convention on Artificial General Intelligence (AGI)—not narrow AI we have today, but general AI we could have within 10 years. If we do not get the initial conditions, rules, and guardrails right for this future AGI, then it could evolve into a superintelligence beyond our control and benefit. This is what Hawking, Gates, and others have warned about.

Some think that such discussions of regulation are premature or that they will hinder the development of AGI, but this overlooks the fact that it could take 10 or more years to create agreements on initial conditions for AGI, then a UN Convention on AGI, and then establish

global governance system. We have to create a global agreement on how to govern AGI BEFORE it is created. This could be the most difficult management challenge we face today.

Exploring synergic relations in our universities instead of only zero-sum thinking should further human security. The US-China Climate Change and AGI synergies are a good place to start.

Even schools of business could contribute to this global mindset change. University Schools of Business teach competitive intelligence, competitive advantage, and competitive strategies. They could also teach potential synergetic intelligence, synergetic advantage, and synergetic strategy.

## 2. Some Strategic Threats to Human Security

United Nations Secretary-General António Guterres included “existential” risks or threats five times in his report *Our Common Agenda*. This UN report also calls for many UN reforms, including a periodic report: Strategic Foresight and Global Risk Report on a regular basis. In response to an informal request to the author by the Executive Office of the UN Secretary-General for very brief overviews of some existential threats to be considered, the following was submitted:

### 2.1. Loss of control over future forms of Artificial Intelligence

As explained above, if the initial conditions of Artificial General Intelligence (AGI) are not “right,” it could evolve into the kind of Artificial Super Intelligence (ASI) that Stephen Hawking, Elon Musk, and Bill Gates have warned could threaten the future of humanity. Intense pressures of competition among corporations and zero-sum power politics among states for advanced AGI could lead to inadequate initial conditions, cutting corners, and other reckless behavior. Instead, synergies among the US and China could lead to a more rational development of AGI and a global governance system.

### 2.2. Massive Discharges of Hydrogen Sulfide (H<sub>2</sub>S) from De-oxygenated Oceans, caused by Advanced Global Warming

Global warming is beginning to change ocean currents. If this trend continues, water conveyors that bring oxygen to the bottom of the ocean will stop. Microorganisms that proliferate without oxygen emit hydrogen sulfide (H<sub>2</sub>S – a deadly gas) when they die. This, plus ozone depletion, may have killed 97% of life during the Permian extinction\*. Also in our future could be desperate attempts at geoengineering that go astray. Again, the synergic strategy could make a difference to human security.

### 2.3. Weakening of the Earth’s Magnetic Shield that Protects us from Deadly Solar Radiation

The Earth’s magnetic fields weaken as the magnetic poles reverse. The last reversal was 42 million years ago†, and scientists predict the Earth is due for another one. The process

\* <https://www.sciencedirect.com/science/article/abs/pii/S0921818112001452>

† <https://www.sciencenews.org/article/earth-magnetic-field-reversal-mass-extinctions-environment-crisis>

of reversal can take hundreds of years, during which time humanity and all life will be vulnerable to deadly radiation\* worldwide. If a solar eruption the size of the 1859 Carrington Event occurs again it would knock out the Internet, electrical systems, water controls, and crucial satellites. If it occurs during a magnetic reversal, it could kill life on Earth.

---

*“We have no rules, agreements, conventions, and governance systems in place to address what Stephen Hawking, Elon Musk, and Bill Gates have warned the public could threaten the future of humanity via the future globally connected Internet of Things (IoT).”*

---

#### **2.4. Malicious Nanotechnology (including the “gray goo” problem)**

There are two approaches to nanotechnology: big machines, making nanotech that we have today, and atomically precise manufacturing and self-assembly that we do not have yet. Theoretically, the second version could take CO<sub>2</sub> from the air, strip out the oxygen, and make massive carbon nanotech structures, with nothing to stop it. This uncontrolled self-assembly is referred to as the “gray goo problem.”†

#### **2.5. A single individual acting alone, could one day create and deploy a weapon of Mass Destruction (most likely from synthetic biology)**

Synthetic biology that mixes genetic material from different species could make a new kind of virus living outside the body for deployment around the world, with a long incubation period. National technical means can identify and disrupt such actions, but probably not all. Improving applications of cognitive science and child development psychology could reduce such insane people, but not all. Families and communities can also help reduce the number of such mass killers. Technologies will continue to become more powerful, decentralized, and easier to use‡, so strategies to prevent misuses should increase—globally—as well.

#### **2.6. Nuclear War Escalation**

Although nuclear war was prevented between the USSR and the USA, the number of countries with nuclear weapons has grown to nine: United States, Russia, France, China, the United Kingdom, Pakistan, India, Israel, and North Korea. Since there are political tensions among several of these the possibility of war is not zero. In addition to deadly radiation, Carl Sagan§ and other scientists explained that firestorms created by the nuclear explosions would fill the atmosphere with sufficient smoke, soot, and dust circling the globe interrupt plant photosynthesis stopping food supply.

\* <https://www.sciencedirect.com/science/article/abs/pii/S1342937X16000319>

† <https://iopscience.iop.org/article/10.1088/0957-4484/15/8/001>

‡ <https://books.google.co.in/books?id=bUbiDgAAOBAJ&pg=PA161&lpg=PA161&dq=#v=onepage&q&f=false>

§ <https://www.atomicarchive.com/science/effects/nuclear-winter.html>

## **2.7. New Uncontrollable, more severe Pandemics**

As synthetic biological research advances and proliferates, the ability to create (by accident or design) immune pathogens that continually mutate increases the possibility, although remote, of human extinction. Human-caused environmental changes could also lead to pathogens that could also lead to our extinction\*. While no single pandemic is likely to extinguish humanity, they may do so in combination with other catastrophic threats.

## **2.8. Particle Accelerator Accident**

Some scientists consider it possible that future particle accelerator experiments could possibly† destroy the Earth and even open a blackhole or create a phase transition that could tear the fabric of space. Brookhaven National Laboratory‡ altered its research program when they found an extremely unlikely chance of opening a blackhole, but they determined the possibility was not zero.

## **2.9. Gamma-ray Bursts**

When two stars collide§, a gamma-ray burst originating thousands of light years away, could sufficiently damage the protective ozone layer to kill life on Earth. According to Dr. Adrian Melott of the Department of Physics and Astronomy at the University of Kansas, “We don’t know exactly when one came, but we’re rather sure it did¶ come—and left its mark.” The WR 104-star system could cause such a gamma-ray burst in the future. October 2022 one just affected Earth’s lightening from 2 billion light years away. The Sun could also emit high-energy flares, damaging our ozone layer.

## **2.10. An Asteroid Collision**

An asteroid large enough to end humanity missed the Earth by six hours on March 23, 1989. If it would have hit the Earth, the impact would have been the equivalent of a thousand of our most powerful nuclear bombs. NASA is identifying and tracking such threats now. There are over 12,000 asteroids 140 meters or more in diameter that pass near Earth’s orbit that could destroy an average sized country. Although some have proposed attacking an asteroid with an explosive device, that could result in multiple hits on the Earth. Instead, research to find effective ways to change its course may prove safer.

Two other existential human security threats could be added: Super Volcano and Extraterrestrial Contact, but the most immediate to address is AGI.

## **3. Why Focus on AGI Now?**

Because it is the most near-term potential existential human security threat. AI is advancing so rapidly, that some experts believe that artificial general intelligence (AGI)

\* <https://www.news-medical.net/news/20210419/Humans-versus-viruses-Can-we-avoid-extinction-in-near-future.aspx>

† <https://www.sciencealert.com/earth-could-be-crushed-to-the-size-of-a-soccer-field-by-particle-accelerator-experiments-says-astronomer>

‡ <https://www.scientificamerican.com/article/the-safety-of-strangelets/>

§ <https://www.space.com/13221-space-collisions-earth-extinctions-gamma-ray-bursts.html>

¶ [https://www.nasa.gov/vision/universe/starsgalaxies/gammaray\\_extinction.html](https://www.nasa.gov/vision/universe/starsgalaxies/gammaray_extinction.html)

could occur before the end of this decade. We have no rules, agreements, conventions, and governance systems in place to address what

Stephen Hawking, Elon Musk, and Bill Gates have warned the public could threaten the future of humanity via the future globally connected Internet of Things (IoT).

There are many excellent centers studying values and the ethical issues of ANI, but not potential global governance models for the transition to AGI. The distinctions among ANI, AGI, and ASI are usually missing in these studies. The U.S. National Security Commission on Artificial Intelligence report has little mention of these distinctions and the US S&T Artificial Intelligence & Machine Learning Strategic Plan has no mention at all. Current work on AI governance is designed to catch up with the artificial narrow intelligence proliferating worldwide today.\*

Meanwhile advances toward AGI seem to be accelerating. Investments into AGI development are forecast to be \$50 billion by 2023<sup>†</sup>. However, estimates on financial investments into AGI are difficult to measure, since government classified funding into AGI is unknown. Microsoft invested \$10 billion into OpenAI according to Bloomberg.<sup>‡</sup> A survey in 2020, found 72 projects working on AGI development in 37 countries.<sup>§</sup>

Although expert judgments vary about when AGI will be possible, the estimates keep coming closer and closer.<sup>¶</sup> Estimates also vary due to definitions. Many say it is human level intelligence or capacity. However, there are many forms of ANI today that are already beyond human capacity or human level intelligence such as:

- Protein folding: AlphaFold by DeepMind;
- Lip reading: LipNet by DeepMind
- Playing games: Chess: Deep Blue by IBM; and Jeopardy & Go: AlphaGo and AlphaZero
- Live voice translation: Microsoft
- Mathematics
- Flying planes, driving trucks
- Face recognition
- Medical diagnosis
- Reading comprehension speed: Microsoft and Alibaba
- Legal analysis: LawGeex
- Income tax preparation: TurboTax
- Organizing shipping: Amazon
- Specific research: Google; Alexa

\* The author is a member of the IEEE SA P2863 Organizational Governance of Artificial Intelligence Working Group

† *Artificial General Intelligence 2018–2023* <https://mindcommerce.com/reports/artificial-general-intelligence-2018-2023/> and <https://www.analyticsinsight.net/microsoft-baidu-tech-giants-head-head-becoming-agi-innovator/>

‡ Microsoft Invests \$10 Billion into ChatGPT maker OpenAI <https://www.bloomberg.com/news/articles/2023-01-23/microsoft-makes-multibillion-dollar-investment-in-openai?leadSource=uverify%20wall>

§ 2020 Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy [https://gerinstitute.org/papers/055\\_agi-2020.pdf](https://gerinstitute.org/papers/055_agi-2020.pdf)

¶ AI Multiple, 995 experts' opinion: AGI / singularity by 2060 [2021 update] <https://research.aimultiple.com/artificial-general-intelligence-singularity-timing/, December 31, 2020>

- Traffic navigation: Google Maps
- AI/robots for repetitive tasks
- Large scale data analysis
- Autonomous vehicles

For the purpose of this paper, AGI is defined as a general-purpose AI that can learn, edit its code, act autonomously to address novel and complex problems with novel and complex strategies similar to or better than humans, as distinct from Artificial Narrow Intelligence (ANI) that has a narrower purpose. Artificial Super Intelligence is AGI that has become independent of humans, developing its own purposes, goals, and strategies without human understanding, awareness, or control and continually increasing its intelligence beyond humanity as-a-whole.

Granted there are grey areas between narrow and general. Large platforms are being created of many ANIs, such as Gato\* by DeepMind of Alphabet which is a deep neural network that can perform 604 different tasks, from managing a robot to recognizing images and playing games—it is not AGI, but Gato is more than the usual ANI: “The same network with the same weights can play Atari, caption images, chat, stack blocks with a real robot arm and much more, deciding based on its context whether to output text, joint torques, button presses, or other tokens.”† Additionally, Wu Dao 2.0 by the Beijing Academy of Artificial Intelligence‡ has 1.75 trillion parameters§ trained from both text and graphic data. This allows it to generate new text and images on command and has its virtual student (Hua Zhibing) that learns from Wu Dao 2.0.¶

AGI should not be confused with General Purpose AI Systems (GPAIS)\*\* which is defined as an AI system “able to perform generally applicable functions such as image/speech recognition, audio/video generation, pattern detection, question answering, translation etc. These systems rely on “transfer learning” applying knowledge from one task to another. ChatGPT†† is an upgrade from GPT-3 to GPT-3.5 that can generate human-like text and perform a wide range of language tasks such as translation, summarization, and question answering. (GPT-3 uses 175 billion machine learning parameters.) ChatGPT interacts with the user to produce sophisticated text from simple instructions or questions. See the Appendix for an example of how it answered the first question in the second section below. It can also write and correct code, write music in different styles, organize information, and other uses being invented now. SingularityNet is also in this grey area. It brings together AI developers who want to create AGI and share code so that AGI might emerge from many interactions. The Athens Roundtable held at the European Parliament on 1-2 December 2022 did discuss

\* <https://www.deepmind.com/publications/a-generalist-agent>

† Overview AI values, principle, an ethics <https://openreview.net/forum?id=1ikK0kHjvj>

‡ Beijing Academy of Artificial Intelligence <https://www.baai.ac.cn/english.html>

§ Beijing-funded AI language model tops Google and OpenAI in raw numbers <https://www.scmp.com/tech/tech-war/article/3135764/us-china-tech-war-beijing-funded-ai-researchers-surpass-google-and>

¶ China unveils first domestically developed virtual student <http://en.people.cn/n3/2021/0604/c90000-9857985.html>

\*\* Council of the European Union General Purpose AI Systems (GPAIS) <https://data.consilium.europa.eu/doc/document/ST-14278-2021-INIT/en/pdf>

†† ChatGPT: Optimizing Language Models for Dialogue <https://openai.com/blog/chatgpt/>

General Purpose AI, but not AGI. The Future of Life Institute has assessed General Purpose AI and the AI Act,\* but not AGI.

It is the business of futurists to explore a range of possible futures. Since some AGI experts believe it is possible to have AGI in just a few years, then that possibility should be taken seriously today. If beneficial AGI's initial conditions are important for creating of AGI that is less likely to evolve into an artificial superintelligence that becomes an existential human security threat, then identification of such initial conditions and global governance systems also should also be taken seriously. Here is some initial thinking on this:

### 3.1. Some Initial Conditions for AGI:

- Regulatory standards in place prior to an AGI being connected to the Internet
- Incentives to cooperate with humans and other AGIs
- Seeks synergies with other AGIs rather than conflicts but notifies humans if a conflict begins.
- Keep detailed records of your design processes and decision making.
- Ability to distinguish between how we act vs. how we should act.
- Ben: We should build a neural-symbolic-evolutionary AGI with rich self-reflective and compassionate capability, educate it well, work with it on benefit projects, put it under decentralized control, and have some of us fuse with it.
- Heuristic Imperatives (reduce suffering, increase prosperity and understanding – David Shapiro)
- Ability to teach itself reality – AlphaZero teaches itself how to win with just the rules as a given – what are the rules for learning reality?
- Criteria to know when to be autonomous and when to check with humans
- Think about Chaos Theory's position in sensitivity to initial conditions as leading indicator of Chaos (meaning behaviour does not match past perceived rules).
- A pause command for the AGI that traces back to see how/who/when the AGI made the decision that led to the undesirable action, and then can be amended (patched?), in conversation (?) with a human. But such patches could build up overtime, creating their own anomalies.
- Transfer learning elements should be pre-audit approved before being added to an AGI and unsupervised learning.

### 3.2. Examples of Rules:

- An algorithm cannot turn off its own off switch<sup>†</sup> or learn how to prevent human intervention (but there are likely to be cases where this would not be desirable – how to address?)

\* *General Purpose AI and the AI Act*, an assessment by the Future of Life Institute <https://artificialintelligenceact.eu/wp-content/uploads/2022/05/General-Purpose-AI-and-the-AI-Act.pdf>

<sup>†</sup> <https://youtu.be/lXSLPwigvi0>



- Cannot use subliminal techniques to manipulate humans.
- Continuous audit system able to pause an AGI that triggers evaluation when an AGI does the unexpected, undesirable action, not anticipated in the utility function, to determine why and how it failed or caused harm.
- Recursive self-improvement and self-replication with human supervision
- Incorporate principles of the Global Partnership on AI, OECD, UNESCO
- Meets IEEE and ISO governance and transparency standards (definitions, principles measurements, and auditing methods).
- IEEE Ethically Aligned Design 15 standards \*
- IEEE SA P2863 (Jglenn member) standard doc ready by 6/2023
- ISO/IEC JTC 1, Information Technology, Subcommittee SC 42, Artificial Intelligence
- Asimov's three laws
- Russel's three Human-Compatible AI principles; Stuart Russel includes uncertainty of what is right, so it can be developed.
- 2017 Asilomar conference – Ethics and Values; e.g., 16 Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives.
- Decision/data logging to recreate a decision and date at the time of an error. (a government transparency law could require X-year period of data retention).
- Reinforces human development rather than the commoditization of individuals.
- Ability to state why the action is requested and the environment in which the mission is to be conducted –Karl Schroeder
- Similar to a flight recorder, the AGI should have a log of changes of the neural network (not all activity).

### **3.3. Some Governance Issues**

- Create with diversity of input—government, business, universities, NGOs, UN agencies, software engineers, poets, futurists, international lawyers
- Trusted enforcement mechanisms
- Global governance agency should have access to all code to review ethics? And how to protect the IP of the coder/corporation?

### **3.4. Audit – Certification – License**

- Tested in several environments (including wild card interventions) to see if its values/principles hold up, and if so, then certified.
- Massively complex simulation used to test software and alignment with stated values with their definitions and measures for audits.
- Continuous audit system to monitor crossing of guardrails, rules infractions, and unethical or biased behaviour.

---

\* <https://ethicsinaction.ieee.org/>

- Show detailed records of design processes and decision making.
- Trust Label if it meets standards and accountability.

### 3.5. Development Strategies

- Can an AI be designed to solve the alignment problems?
- MIRI: break down the alignment problem into simpler and more precisely stated sub-problems, develop basic mathematical theory for understanding these problems, and then make use of our newfound understanding in engineering applications.
- Create and use less powerful versions of AGI (e.g., future versions of GPT, etc.) one after the other to learn how to manage AGI and prevent “one shot to get it right” future.
- Common platforms for AGI developers and their cryptocurrencies
- Pause development beyond GPT-4 to assess situation
- Causal reasoning based on a conceptual model of reality (explore, as a global ontology of sufficient amount of the world). Peer reviewed cognitive maps of how the world works—and use of massively complex simulation of global human behavior and natural environment.
- Evidence which now strongly supports the claim of predictive learning hard wired in the mammal brain\*. Predictive learning in engineering is VERY closely tied to Kalman filtering and state estimation... building up an image of the state of the world. Adaptive engineering systems WITHOUT that capability scale poorly.

## 4. Initial Sample of Potential Governance Models for AGI

These are drawn from “Artificial General Intelligence Issues and Opportunities,” by Jerome C. Glenn contracted by the European Commission for input to Horizons 2024-27 planning.

1. IAEA-like model or WTO-like with enforcement powers. These are the easiest to understand, but likely to be too static to manage AGI.
2. IPCC-like model in concert with international treaties. This approach has not led to a governance system for climate change.
3. Online real-time global collective intelligence system with audit and licensing status, governance by information power. This would be useful to help select and use an AGI system, but no proof that information power would be sufficient to govern the evolution of AGI.
4. GGCC (Global Governance Coordinating Committees) would be flexible and enforced by national sanctions, ad hoc legal rulings in different countries, and insurance premiums. This has too many ways for AGI developers to avoid meeting standards.
5. UN, ISO and/or IEEE standards used for auditing and licensing. Licensing would affect purchases and would have impact, but requires international agreement or treaty with all countries ratifying.

---

\* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9262088/>

6. Put different parts of AGI governance under different bodies like ITU, WTO, WIPO. Some of this is likely to happen but would not be sufficient to govern all instances of AGI systems.
7. Decentralized Semi-Autonomous TransInstitution. This could be the most effective, but the most difficult to establish since both Decentralized Semi-Autonomous Organizations and TransInstitutions are new concepts.

Clearly there are many threats to human security, and all should be addressed, but one of the greatest opportunities to improve human security is the transition from zero-sum to synergetic thinking and applying that to a US-China United Nations Convention on AGI.

*Author's Contact Information*

*Email: [jerome.glenn@millennium-project.org](mailto:jerome.glenn@millennium-project.org)*