



FORESIGHT ON DEMAND: “FORESIGHT TOWARDS THE 2ND STRATEGIC PLAN FOR HORIZON EUROPE”

Artificial General Intelligence Issues and Opportunities

RAPID EXPLORATION

Author: Jerome Glenn (The Millennium Project)

February 2023



CONTENT

Summary	4
1. Introduction.....	5
2. Current situation	5
2.1. Driving Forces.....	6
2.2. Gray area between Narrow and General – General Purpose AI.....	7
2.3. Some Current Questions About Future AGI.....	7
2.4. Relationship between quantum computing and AGI	8
2.5. Some Examples of Potential Beneficial Initial Conditions for AGI.....	9
3. Distilled Alternative views on the Evolution of ANI to AGI to ASI.....	9
3.1. AGI Evolution in Parallel with Humanity	9
3.2. AGI Merger with Humanity.....	10
3.3. Mixed Complexity, Backing into Agreements, and Science Fiction Nightmares	10
4. AGI Issues for Research & Innovation Policy	11
4.1. Initial Conditions for AGI	11
4.2. Centralization vs. Decentralization	11
4.3. Are Regulations Blocking Development of AGI a Valid Concern?	11
4.4. Ownership and Liability of AGI and Its Products	11
4.5. Anticipating employment and other economic impacts	12
4.6. How to Counter Organized Crime’s use of AGI?.....	12
4.7. Information Warfare	12
4.8. AGI Governance	13
4.9. Some potential governance models	13
Concluding Recommendations.....	14
Annex.....	15

SUMMARY

National governments and multi-lateral organizations like the European Union, OECD, and UNESCO have identified values and principles for artificial narrow intelligence and national strategies for its development. But little attention has been given to identifying the beneficial initial conditions for future Artificial General Intelligence (AGI). The initial conditions for AGI will determine if Artificial Super Intelligence will evolve to benefit humanity or not. Even if international agreements are reached on the beneficial initial conditions for AGI, a global governance system will still be needed to enforce them and oversee the development and management of AGI. Since it may take ten, twenty, or more years to create and ratify an international AGI treaty and establish a global AGI governance system, and since some experts believe it is possible to have AGI within ten to twenty years, it is therefore important to work on these issues as soon as possible. The most critical AGI issues are its initial conditions and global governance. These issues are important for governments to get right from the outset.

1. INTRODUCTION

AI is advancing so rapidly, that some experts believe that artificial general intelligence (AGI) could occur before the end of this decade;¹ hence, it is time to begin serious deliberations about AGI. There are three categories of Artificial Intelligence (AI): narrow, general, and super.

Artificial Narrow Intelligence (ANI) sometimes called weak AI is the kind of AI or machine learning that we have today: each software application has a single specific purpose. However, generalist agents are being created that can perform several functions, but not as general and creative as AGI.

Artificial General Intelligence (AGI) sometimes called strong AI is similar to human capacity for novel problem-solving and reasoning whose goals are set by humans. It can: address complex problems without pre-programming like ANI requires; initiate searches for information worldwide; use sensors and the Internet of Things (IoT) to learn; make phone calls and interview people; make logical deductions; re-write or edit its code to be more intelligent ...continually, so it gets smarter and smarter, faster and faster than humans. Some believe this could happen within ten years;² some others argue that AGI is impossible for many more years, if ever.³ Although there is no consensus within the AI community, some would say AGI will have a unique form of sentience.

Artificial Super Intelligence (ASI) is AGI that becomes so advanced that it sets its own goals and strategies independent of human awareness or understanding. It is most likely to emerge from AGI.⁴ It is unknown how fast ASI could emerge from AGI. It could be almost immediately, or years, or never. Hence, research & innovation policy should consider a range of possibilities. Allan Dafoe, of DeepMind and the Future of Humanity Institute, University of Oxford, says that “the governance of AI is the most important issue facing humanity today ...”⁵ Elon Musk believes that the single, most pressing existential issue we face is how the advent of ASI is or is not symbiotic with humanity.⁶

2. CURRENT SITUATION

AGI does not exist. President Putin has said whoever leads AI will rule the world. China plans to lead international competition by 2030.⁷ Although “AGI” or “Artificial General Intelligence” does not appear in the State Council Notice on the Issuance of the Next Generation Artificial Intelligence Development Plan: A Next Generation Artificial Intelligence Development Plan (Released: July 20, 2017), terms such as “strong generalization capabilities...AI key general technology system...cross-medium analytical reasoning technology” does seem like AGI and the plan states that China will be “occupying the commanding heights of AI technology.” Since ANI is with us now, one can assume that President Putin and the Chinese Plan were referring to AGI. Therefore, it is also reasonable to assume that the “Great AGI Race” is on with both governments and corporations. In such a race, Deep Mind Co-founder and CEO Demis Hassabis said people may cut corners making future AGI less safe. Adding to this race are the Brain Projects⁸ in the EU, USA, China and Japan, and other neuroscience advances.

¹ *When will singularity happy? 995 experts' options on AGI* (updated September 26, 2022), <https://research.aimultiple.com/artificial-general-intelligence-singularity-timing/>

² *Artificial General Intelligence: The Future Of Decision AI* <https://datatechvibe.com/ai/artificial-general-intelligence-the-future-of-decision-ai/>

³ *Why Machines Will Never Rule the World Artificial Intelligence without Fear* By Jobst Landgrebe, Barry Smith <https://www.routledge.com/Why-Machines-Will-Never-Rule-the-World-Artificial-Intelligence-without/Landgrebe-Smith/p/book/9781032309934>

⁴ ASI could evolve in parallel to AGI's development or sui generis, but this paper assumes it would emerge from AGI.

⁵ Video of talk by Allan Dafoe <https://youtu.be/ug6X67xU7Kg>

⁶ *Elon Musk: 'Mark my words — A.I. is far more dangerous than nukes'* <https://www.cnn.com/2018/03/13/elon-musk-at-sxsw-a-i-is-more-dangerous-than-nuclear-weapons.html>

⁷ State Council Notice on the Issuance of the Next Generation Artificial Intelligence Development Plan, released: July 20, 2017 <https://d1y8sb8igg2f8e.cloudfront.net/documents/translation-fulltext-8.1.17.pdf>

⁸ Inventory of Brain Projects Working Group <https://www.internationalbraininitiative.org/inventory-brain-projects-working-group>

Estimates on financial investments into AGI development vary widely. Government classified funding into AGI is unknown. Microsoft alone invested \$1 billion into OpenAI during 2021. And one estimate between 2018 and 2023⁹ projects a total of \$50 billion. A survey in 2020, found 72 projects working on AGI development in 37 countries.¹⁰ Expert judgments about when AGI will be possible also vary broadly. Some working to develop AGI believe it is possible within ten years.^{11,12} Others think it is impossible to ever be achieved. Definitions of AGI also vary. 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 Deep Mind
- 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
- Traffic navigation: Google Maps
- AI/robots for repetitive tasks
- Large scale data analysis
- Autonomous vehicle

One way to understand the distinction between ANI and AGI is an analogy with the human body. Just as the autonomous nervous system includes many narrow purpose systems (respiratory, skeletal, digestive, etc.) that manages much of our body, freeing the mind to be creative, so too many ANI systems are being put in place to run much of the urban infrastructure, which will free much of humanity toward a self-actualization economy.¹³

Usually, large amounts of data are used to make the ANI agent smart. Although DeepMind used volumes of data from expert games to train Alpha Go to defeat world champions, it developed AlphaZero with no data input other than the rules of Go. In 2017, AlphaZero learned by playing against a copy of itself, and in just three days was able to beat AlphaGo.¹⁴

The work on ANI values is well underway around the world and progress is being made on standards and regulations, but not for AGI. The International Organization for Standardization and the Electrotechnical Commission - ISO/IEC Joint Technical Committee for Information Technology created a standards sub-committee for Artificial Intelligence (SC 42) jointly lead by the US and China. The ISO standards have been released and the IEEE will do so in a few months (the author is a voting member of the IEEE P2863 AI governance standards working group) but these too just look at ANI.

2.1. Driving Forces

1. Political power – Putin’s statement, China’s 2030 goal, zero-sum geopolitical power motivation.
2. Advances in cognitive science that could be applied to algorithms (Human brain projects in the US, EC, China, and Japan).

⁹ *Investing in Artificial General Intelligence or AGI* <https://www.nanalyze.com/2020/12/investing-artificial-general-intelligence-agi/> and *From Microsoft to Baidu, Tech Giants Head-to-Head for Becoming AGI Innovator* <https://www.analyticsinsight.net/microsoft-baidu-tech-giants-head-head-becoming-agi-innovator/>

¹⁰ *2020 Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy* https://gcrinstitute.org/papers/055_agi-2020.pdf

¹¹ *AI Multiple, 995 experts’ opinion: AGI / singularity by 2060 [2021 update]* <https://research.aimultiple.com/artificial-general-intelligence-singularity-timing/>, December 31, 2020

¹² *Date Weakly General AI is Publicly Known* <https://www.metaculus.com/questions/3479/date-weakly-general-ai-is-publicly-known/>

¹³ Scenario 3: If Humans were free: the Self-Actualization Economy in Work/Technology 2050: Scenarios and Actions, The Millennium Project, Washington, DC 2020

¹⁴ *AlphaGo Zero: Starting from scratch* <https://www.deepmind.com/blog/alphago-zero-starting-from-scratch>

3. Increasing funding allocated to AGI research in corporations; e.g., Demis Hassabis with DeepMind; OpenAI lead by Sam Altman; Facebook's FAIR with Yann LeCun (one of the fathers of deep learning); Alibaba has Alibaba A.I. Labs, Baidu has Baidu Research; Tencent has the Tencent AI Lab; and the Beijing Academy of Artificial Intelligence has Wu Dao.
4. Increases in computer capacity including progress with quantum computing.
5. Increasing volume and organization of data.
6. Collaborative, synergetic efforts like SingularityNet sharing and selling algorithms and code and potential for transfer learning.
7. Increasing democratization of AI tools and education.
8. The human desires for achievement, excitement, and having greater impact with less effort.

2.2. Gray area between Narrow and General – General Purpose AI

Large platforms are being created of many ANIs such as Gato¹⁵ 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.”¹⁶ Also the Wu Dao 2.0 by the Beijing Academy of Artificial Intelligence¹⁷ has 1.75 trillion parameters¹⁸ 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 WuDwo 2.0.¹⁹

AGI should not be confused with General Purpose AI Systems (GPAIS)²⁰ 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. ChatGPT²¹ 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 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 gray area. It brings together AI developers who want to create AGI and share code such that AGI might emerge from many interactions. The Athens Roundtable held at the European Parliament on 1-2 December 2022 did discuss General Purpose AI, but not AGI. The Future of Life Institute has assessed General Purpose AI and the AI Act,²² but not AGI.

2.3. Some Current Questions About Future AGI

AGI is expected to address novel and extremely complex problems by initiating research strategies from exploring the Internet of Things (IoT), interviewing experts, making logical deductions, learning from experience and reinforcement without the need for its own massive databases, and continually editing and re-writing its own code to continually improve its own intelligence. An AGI might be tasked to create plans and strategies

¹⁵ <https://www.deepmind.com/publications/a-generalist-agent>

¹⁶ Overview AI values, principle, an ethics <https://openreview.net/forum?id=1ikK0kHivj>

¹⁷ Beijing Academy of Artificial Intelligence <https://www.baai.ac.cn/english.html>

¹⁸ 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>

¹⁹ China unveils first domestically developed virtual student <http://en.people.cn/n3/2021/0604/c90000-9857985.html>

²⁰ Council of the European Union General Purpose AI Systems (GPAIS) <https://data.consilium.europa.eu/doc/document/ST-14278-2021-INIT/en/pdf>

²¹ ChatGPT: Optimizing Language Models for Dialogue <https://openai.com/blog/chatgpt/>

²² 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>

to avoid war, protect democracy and human rights, manage complex urban infrastructures, meet climate change goals, counter transnational organized crime, and manage water-energy-food availability.

To achieve such abilities without the future nightmares of science fiction, global agreements with all relevant countries and corporations will be needed. To achieve such an agreement or set of agreements, many questions should be addressed. For example:

- How to manage the international cooperation necessary to build international agreements and a governance system while nations and corporations are in an intellectual “arms race” for global leadership. (IAEA and nuclear weapon treaties did create governance systems during the Cold War arms race.)
- And related: How can international agreements and a governance system prevent an AGI “arms race” and escalation from going faster than expected, getting out of control and leading to war, be it kinetic, algorithmic, cyber, or information warfare?
- How can international treaties and a governance system prevent increased centralization of power crowding out others?
- If IAEA, ITU, WTO, and other international governance bodies were created today, how would officials of such agencies create them differently, and what advice might they give for creating uniquely complex and dynamic AGI governance?
- Drawing on the work of the Global Partnership on Artificial Intelligence (GAPI) and others that have already identified norms, principles, and values, what will be the most acceptable combination or hierarchy of values as the basis for international treaties and a governance system?
- How can a governance model help assure an AGI is aligned with acceptable global values, priorities among values, and accommodate future changes or additions to values?
- How can a governance model correct undesirable action unanticipated in utility functions?
- How to develop and enforce AGI algorithm audit standards?
- How can the use of ANI-to-AGI by organized crime and terrorism be reduced or prevented?
- To what degree do AGI thought leaders and primary stakeholders agree about the framing of governance issues?
- Should an international governance trial, test, or experiment be constructed first with a single focus (e.g., health or climate change), and then to learn the rules and standards from such experiences to extend to broader governance of the transition from ANI to AGI?
- Should AGI have rights? If so, which ones?
- How can governance be flexible enough to respond to new issues previously unknown at the time of creating that governance system and old issues reinterpreted in the future?

Since the EC has led on some complex multilateral agreements, it could perform a great service by addressing some of these questions.

2.4. Relationship between quantum computing and AGI

Quantum computing could speed-up the arrival of AGI by improving and running machine learning algorithms.²³ Quantum computing is progressing rapidly and likely to be quite advanced when AGI arrives: Both [IBM](#) and [Google](#) announced they plan to have a 1-million-qubit chip by 2030. Assuming massive engineering problems can be solved as the number of qubits increases, it is expected that quantum computing will be able to create very large, complex simulations. Over the next ten to twenty years the nature of the simulations produced by future quantum computing might evolve to be able to test AGIs alignment with stated values in many more situations than currently available today. If so, this could be one future approach to auditing and certifying an AGI for a license prior to public use.

²³ *The quantum boost to AI paves the way for AGI* <https://dataconomy.com/2022/06/quantum-artificial-intelligence/>

2.5. Some Examples of Potential Beneficial Initial Conditions for AGI

Since the initial conditions for AGI will determine or at least influence how beneficial ASI will be for humanity, special attention should be given to identifying, assessing, and as possible, testing these conditions. The following list is offered to simulate discussion to more rigorously create initial conditions that might be more carefully defined with metrics suitable for audits, and that could be used for future licensing possibilities.

- Values agreed to for ANI²⁴ e.g., ISO, IEEE, and OECD²⁵.
- Massively complex simulations to test AGI's alignment with values.
- Able to be continuously audited.
- Contain an automatic off-switch (or a signal to Interpol for automatic closure) when a failure during the continual audit is detected.
- Ability to distinguish between how we act vs. how we should act.
- Ability for humans to check why the AGI is doing something questionable or undesirable. A pause command 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?), interacting with a human. However, such patches could build up overtime, creating their own anomalies; this has to be considered.
- Self-replication with human supervision.
- Cannot turn off its own off-switch. An over-ride switch - allows AGI to be switched off by a human.
- Explain how it is possible to build AGI that is more intelligent than humans but does not have more power than humans?
- Transfer learning elements should be pre-audit approved before added to an AGI and unsupervised learning.
- Guard rails to catch unanticipated behavior and side effects that triggers analysis to determine cause.
- Able to stop an AGI that seems to be causing a conflict that could create a war.
- Seeks synergies with other AGIs rather than conflicts but notifies humans if a conflict begins.
- Cannot use subliminal techniques to manipulate humans for nefarious purposes.
- AGI developers should be required to keep detailed records during design and development process accessible by the future governance system.
- Decision/data logging to recreate a decision and date at the time of an error.
- Similar to a flight recorder, the AGI should have a log of changes of the neural network (not all activity).
- Reinforces human development and individual empowerment rather than the commoditization of human beings.

3. DISTILLED ALTERNATIVE VIEWS ON THE EVOLUTION OF ANI TO AGI TO ASI

3.1. AGI Evolution in Parallel with Humanity

In this view, initial conditions for AGI incorporate the best of human values and safe guards for humanity as determined by many working groups such as IEEE, ITU, OECD, UNESCO, the InterAcademy Partnership, and leading corporations working on AGI and university experts. These groups define the standards and metrics for AGI audits and draft text for a multi-lateral agreement. Under the convening power of the United Nation, these groups draft the international agreement(s) for governing AGI with enforcement powers. For example, if an AGI fails an audit but did not freeze the program or an unlicensed AGI was discovered, the Interpol would be authorized by the treaty to immediately freeze the AGI. Since AGI can continuously re-

²⁴ https://wilkins.law.harvard.edu/misc/PrincipledAI_DocumentTimeline.jpg

²⁵ Global Partnership on AI (GPAI) headquartered at OECD has two Centers of Expertise: one in Montréal and one in Paris.

write its code, the auditing process would have to be a continuous process not an annual event like a financial audit.

Humans make ethical and unethical decisions. If AGI's initial conditions include the best of human values, then when AGI takes on autonomous decisioning, its decisionmaking should be more ethical than humans as a generalization. Just as ANI will driver cars more safely than humans in general, AGI could make a more ethical society. Much of the built environment will be run by many different ANIs, which in turn could be managed by AGI.

Just as the US and the USSR found ways to manage nuclear weapons avoiding World War III, the US and China could agree to find ways to collaborate and lead the world to stabilize and reverse climate change²⁶ and manage the development and governance of AGI. It became clear that the rush to get to AGI by governments of USA, China, EU, Japan, Russia, and Israel along with corporations like Open AI, DeepMind, Google Brain, Baidu, Meta AI, Alibaba and others could mean that insufficient attention would be paid to the "right" initial conditions. And in this rush, several AGIs with insufficient initial conditions might interact with other AGIs in destructive ways and allow the emergence of an Artificial Super Intelligence (ASI) in a manner that Hawking, Musk, Gates, and others warned about. As a result, international agreements were reached and a governance system was finally put in place. A unique element of the governance system was the creation of an AGI that is tasked by the UN's International AGI Agency, to continually help improve global collaboration among UN agencies, governments, businesses, and research institutes for continual improvement of the governance system of AGI for the benefit of an ever-evolving civilization on Earth and beyond. In short, only an AGI was able to keep up with the evolution of other AGIs and able to help govern them.

3.2. AGI Merger with Humanity

This is similar to 2.1 AGI Evolution in Parallel with Humanity except that it develops AGI while ANI is integrated in or on the human body. Neuralink²⁷ has created computer extensions thin enough to be inserted next to indivial neurons in the brain and record activity.²⁸ A monkey with the implant has played the electronic game pong, with thought alone. The goal is to merge future AI with the brain to so that AGI and the human can evolve together. "If you can't beat them, join them." The merger of human and AGI seems the preference of Chinese research.²⁹

Another approach to the AI-brain merger or direct brain to computer interface is Synchron with its stentrode device and bluetooth connection that began human trails in November 2020.³⁰ The stentrode electrode was injected in a major blood vessel without the need for brain surgery that allowed the patients to manage email and other computer operations. It is reasonable to assume these and other related technologies will evolve over the next ten to twenty years, intime for AGI and human thought to merge or become a continuum.

In this view, it is not necessary for all or even most of humanity to be AGI-augmented for the co-evolution of human and machine. Just enough for the co-evolution to be expressed and tested in enough situations to make the evolution to ASI from AGI to be beneficial for humanity.

3.3. Mixed Complexity, Backing into Agreements, and Science Fiction Nightmares

The AGI race intensifies among countries and corporations without an international agreement for governance. Different AGIs get created differently in different countries, corporations, and organized crime. To prevent one AGI from taking over the world and the proliferation of organized crime AGIs, international efforts

²⁶ The UN Secretary-General suggested that the US and China lead the world on agreements and programs to address Climate Change during the opening session of the UN COP 27 in Egypt.

²⁷ <https://neuralink.com/>

²⁸ *New Brain Implant Begins Human Trials - Neuralink Update!* Video <https://youtu.be/M5w8s2PKwxI>

²⁹ *China's "New Generation" AI-Brain Project* <https://ndupress.ndu.edu/Media/News/News-Article-View/Article/2846343/chinas-new-generation-ai-brain-project/>

³⁰ <https://synchron.com/technology>

are hastily put together to create a management treaty, but there are no enforcement powers for the governance system. There are also several AGI systems that evolved from peer-to-peer developers' platforms. These became AGI decentralized autonomous organizations owned and controlled by many people rather than corporations or governments. As the different AGIs began to interact with each other, ASI emerges from the totality of human interactions on the web, Internet of Things, and interreacting various AGI software without human awareness, understating, or control.³¹ Things begin happening that cannot be explained and then nightmares of science fiction began to proliferate.

4. AGI ISSUES FOR RESEARCH & INNOVATION POLICY

4.1. Initial Conditions for AGI

This has been explained above as the first step in creating beneficial AGI for humanity. Since it is likely to take ten more years to create AGI, there is time to create the list and revisit it many times. The next step is to define these conditions so precisely that they can be audited by a third party. If the EC will support the development of such a list, it should also include audit companies in the assessment to work with software experts on how to create a continuous audit software attached to the eventual AGI.

4.2. Centralization vs. Decentralization

Some fear that centralization of proprietary AGI research within a few governments and corporations could lead to a hegemonic power, and further concentration of wealth crowding out any further AGI competition. China plans to lead international competition by 2030³² and has an advantage of using its vast data from its social credit system. Since other leading countries and corporations might also like to be the leader, it is reasonable to assume that the race for AGI central control is occurring today.

In response to centralized advanced research, the Decentralized AI Alliance with 58 partners³³ is creating a common platform for the exchange of code among advanced AI developers to avoid re-inventing the wheel or its spokes, and create synergies to improve efficiency and speed toward AGI. Other individual examples of decentralized platforms for AI developers to cooperate on the peer-to-peer development of AGI are Cortex, SingularityNet, Matrix AI, and Open Economic Framework of Fetch.ai. Should the EU support both centralized AGI by the EC, or decentralized development of AGI, or both?

4.3. Are Regulations Blocking Development of AGI a Valid Concern?

It is argued that creating rules for governance of AGI too soon will stifle its development. Since it is likely to take ten or more years to develop AGI international agreements, design an international governance system, and begin implementation, then regulations would not be in place soon enough to block early development work for AGI. This concern over-rates how fast international management can be created. Hence, AGI development would have ten or more years before any regulations would be in place and enforceable.

4.4. Ownership and Liability of AGI and Its Products

Who will own the intellectual property produced by AGI with participation of many inputs from humans, ANI, and sensors around the world? How should liability be decided and who pays for damages caused by AGI? Is it the auditing company that certified the AGI? The government agency that approved the license? The AI programmer that wrote the code? The company that produced, sold, or rented the AGI? The user that defined the purpose, objective, and how the AGI was deployed? UNESCO³⁴ offers an online course to address these issues. If the human "owner" of an AI/Avatar or digital twin dies, and these ANIs or maybe even and

³¹ *I tried Warning them*, video of Elon Musk discussing ANI <https://www.youtube.com/watch?v=9LLmUKG9Toc>

³² State Council Notice on the Issuance of the Next Generation Artificial Intelligence Development Plan, released: July 20, 2017 <https://d1y8sb8igg2f8e.cloudfront.net/documents/translation-fulltext-8.1.17.pdf>

³³ List of partners of the Decentralized AI Alliance <https://daia.foundation/partners/>

³⁴ <https://www.unesco.org/en/articles/unesco-global-mooc-ai-and-rule-law-engaged-thousands-judicial-operators>

AGIs continues to move around the Internet and eventually causes crimes, who is liable? The participants in the Athens Roundtable on AI and the Law³⁵ are grappling with these issues and a good source for exploring these issues further. Although they are exploring law related to ANI not AGI; nevertheless, there should be much overlap on the legal issues.

4.5. Anticipating employment and other economic impacts

Previous technological revolutions from the agricultural age to industrial age and on to the information age created more jobs than each age replaced. But the advent of AGI and its impacts on employment will be different this time because of: 1) the acceleration of technological change; 2) the globalization, interactions, and synergies among NTs (Next Technologies such as synthetic biology, nanotechnology, quantum computing, 3D/4D printing, robots, drones, computational science, as well as both ANI and especially AGI; 3) the existence of a global platform—the Internet—for simultaneous technology transfer with far fewer errors in the transfer; 4) standardization of data bases and protocols; 5) few plateaus or pauses of change allowing time for individuals and cultures to adjust to the changes; 6) billions of empowered people in relatively democratic free markets able to initiate activities; and 7) machines that can learn how you do what you do, and then do it better than you. Anticipating the possible impacts of AGI and preparing for the impacts prior to the advent of AGI could prevent social and political instability,³⁶ as well as facilitate it broader acceptance.

4.6. How to Counter Organized Crime's use of AGI?

Cybercriminal income is about \$1.5 trillion per year.³⁷ Its damages cost \$3 trillion in 2015; \$6 trillion in 2021; and now forecast to be \$25 trillion by 2025.³⁸ Since organized crime has some of the best software experts that money can buy, organized crime could invent and deploy new criminal uses of AGI. Even today, perfectly good contemporary ANI software can be re-purposed for deadly purposes for terrorist and war crimes. For example, MegaSyn by Collaborations Pharmaceuticals designed to create new medical drugs, was tasked to see if it could also come up with toxic substances for bioweapons. In less than six hours over night, it produced a list of 40,000 toxic molecules both known substances and new, unknown substances.³⁹ It is reasonable to assume that criminal use of AGI could pose even greater threats. Work on the initial beneficial AGI conditions should include how to prevent organized crime's use of AGI. The United Nations Interregional Crime and Justice Research Institute has opened the Centre for Artificial Intelligence and Robotics to explore this issue; and hence, the EC should collaborate with this center in research and policy innovation in this issue.

4.7. Information Warfare

Information warfare (as different than cyber warfare that attacks computers, software, and command control systems) manipulates information trusted by targets without their awareness, so that the targets will make decisions against their interest but in the interest of the one conducting information warfare.⁴⁰ ANI is increasingly manipulating perceptions of truth, while the public does not know how to defend itself. "Deep fakes" are increasingly being used with impunity. With AGI, the manipulation of truth is likely to become far worse, but AGI could also be tasked to counter disinformation.

³⁵ <https://www.aiathens.org/>

³⁶ Glenn, Jerome and the Millennium Project team, *Work/Technology 2050: Scenarios and Actions*, The Millennium Project, Washington, DC, 2020.

³⁷ <https://dataprot.net/statistics/cybercrime-statistics/>

³⁸ *Cybercrime to Cost the World \$10.5 Trillion Annually by 2025* <https://cybersecurityventures.com/cybercrime-damages-6-trillion-by-2021/>

³⁹ *AI Drug Discovery Systems Might Be Repurposed to Make Chemical Weapons, Researchers Warn* <https://www.scientificamerican.com/article/ai-drug-discovery-systems-might-be-repurposed-to-make-chemical-weapons-researchers-warn/>

⁴⁰ Glenn, J and Millennium Project Team, *Global Challenge 10, State of the Future 19.1*, The Millennium Project. Washington, DC 2021.

4.8. AGI Governance

Since ASI can evolve from AGI without human awareness and understanding, the only way to influence ASI's relations to humanity—it seems—is a governance system of the initial conditions for AGI. Such a governance system poses unique challenges: 1) it should be dynamic enough to keep up with and even anticipate AGI's changes; 2) it should be enforceable globally, otherwise an AGI could be created without the beneficial initial conditions that evolves eventually to be destructive; and 3) it should be established before AGI exists, otherwise the speed of AGI's evolution would be beyond the ability for regulations and governance to catch up. Each of these seem unfeasible; taken together, it seems impossible.⁴¹ Of course, circumnavigating the globe, human flight and landing on the moon were also thought impossible. Although successful governance of AGI seems impossible today, it has to be pursued, otherwise the evolution to ASI is left to chance and the possibility of alternative nightmare scenarios. Allan Dafoe, of DeepMind and the Future of Humanity Institute, University of Oxford, says that “the governance of AI is the most important issue facing humanity today...”⁴² During the Asilomar conference on Beneficial AI,⁴³ Elon Musk said: “Can government speed keep up with AGI growth speed? ... We need regulation.” and Demis Hassabis, CEO of DeepMind added, “...it's good to start the conversation now” [about AGI governance].

Assuming a governance system would include the requirement that an AGI and its use be licensed, and a license would require an audit, then any audit software would also have to be connected to the AGI so that it would continually be auditing and evolving with the AGI on a continual basis. Other issues that should be addressed in the development of a governance system include: Can an AGI be inspected if it does not want to be inspected, hiding its newly and constantly rewritten code? Can an algorithm be prevented from knowing how it is evaluated? Are off-switches possible in AGI? Specific regulations should be tailored differently for each: software developers, sales companies, buyers, and users. Would state registry trade secrets of AGI's be constantly out of date as it changes? How to verify that an AGI system is being built according the rules prior to an audit? How to ensure that an algorithm should not learn how to prevent human intervention – but there are cases where this would not be desirable – how to address?

4.9. Some potential governance models

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. International S&T Organization (ISTO) as an online real-time global collective intelligence system; 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 of all countries ratifying.
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. Decentralize Semi-Autonomous TransInstitution. This could be the most effective, but the most difficult to establish since both Decentralize Semi-Autonomous Organizations and TransInstitutions are new concepts.

⁴¹ *Superintelligent AI May Be Impossible to Control; That's the Good News Postcard from the 23rd century: Not even possible to know if an AI is superintelligent, much less stop it.* <https://spectrum.ieee.org/super-artificialintelligence>

⁴² Video of talk by Allan Dafoe <https://youtu.be/ug6X67xU7Kq>

⁴³ Video of panel discussion of AGI leaders Elon Musk, Stuart Russell, Ray Kurzweil, Demis Hassabis, Sam Harris, Nick Bostrom, David Chalmers, Bart Selman, and Jaan Tallinn, and Max Tegmark (moderator) <https://youtu.be/h0962biiZa4>

CONCLUDING RECOMMENDATIONS

1. Since it may be difficult to get the major countries and corporations in agreement on an AGI treaty text, the EU could take the leadership to develop the EU's text. Such a text could be used to help speed the agreements among the other nations and corporations for an AGI treaty.
2. Coordinate with other international organizations like the OECD, ITU, WTO, Global Partnership on AI (GPAI), and the UNSG's Office to begin discussions with leading AGI governments, corporations, NGOs, and universities for an AGI treaty.
3. Since the private sector is increasingly involved in the development of AGI, it is likely to occur regardless of government and multi-lateral organizations involvement, but AGI's regulation will not occur without government and multi-lateral involvement. Hence, the EC and other multi-lateral organizations should lead in the development of international agreements on AGI.
4. Issue calls for research to develop the beneficial initial conditions for AGI and governance models to enforce them.
5. Issue calls for how auditing firms could develop unique auditing systems for AGI.
6. Begin discussion with Interpol on enforcement that includes preventing illegal development, use, and sales of AGI.
7. *Manifesto on Enforcing Law in the Age of "Artificial Intelligence:" Transatlantic Reflection Group on Democracy and the Rule of Law in the Age of "Artificial Intelligence"*⁴⁴ recommendations are in line with this paper; however, they address ANI more than AGI; nevertheless, these recommendations should be considered in the development of text for a future AGI treaty.
8. Invest in education and training of EC professionals to manage the transition from ANI to AGI.
9. Consider if the EU should give AI to lower income countries to avoid other countries or corporations making unregulated ANI and AGI available, if international agreements are not yet completed and enforced.

⁴⁴ *A Manifesto on Enforcing Law in the Age of "Artificial Intelligence" Transatlantic Reflection Group on Democracy and the Rule of Law in the Age of "Artificial Intelligence"* 10 October 2022, Rome
<https://static1.squarespace.com/static/604b05672bbe0130d9171569/t/63616fa01b1220098eb291a3/1667329952796/A+Manifesto+on+Enforcing+Law+in+the+Age+of+%E2%80%9CArtificial+Intelligence%E2%80%9D.pdf>

ANNEX

ChatCPT Examples from <https://chat.openai.com/>

ChatCPT was asked the first question in section 1.3 of this paper: How to manage the international cooperation necessary to build international agreements and a governance system while nations and corporations are in an intellectual “arms race” for global leadership. And ChatCPT answered:

Managing international cooperation and building international agreements and governance systems can be challenging, especially when nations and corporations are in competition for global leadership. One approach to addressing this challenge is to focus on creating a shared understanding of the benefits of cooperation and the costs of not working together. This can involve fostering dialogue and communication among nations and corporations, as well as promoting the development of mutually beneficial agreements and policies. Additionally, it may be helpful to establish mechanisms for dispute resolution and conflict management to help prevent and address conflicts that may arise. Overall, the key to managing international cooperation in this context is to find ways to incentivize cooperation and address potential sources of conflict.

A Further question to ChatCPT on this answer was: What are the benefits of international cooperation on the development of artificial general intelligence - AGI? And ChatCPT answered:

“There are several potential benefits of international cooperation on the development of artificial general intelligence (AGI), including:

Shared resources and expertise: International cooperation can help to pool resources and expertise from different nations and corporations, allowing for more rapid and effective development of AGI technologies.

Improved safety and ethical considerations: By working together, nations and corporations can ensure that AGI technologies are developed in a safe and ethical manner. This can involve establishing guidelines and protocols for AGI development, as well as conducting research on the potential risks and benefits of AGI.

Economic benefits: AGI has the potential to drive significant economic growth and innovation. International cooperation can help to ensure that the benefits of AGI are shared widely and that all nations and corporations have access to the latest AGI technologies.

Improved global security: AGI has the potential to impact global security in both positive and negative ways. International cooperation can help to ensure that AGI is used for the common good and to prevent its misuse for malicious purposes.

Overall, international cooperation on the development of AGI has the potential to accelerate progress in this field, promote shared benefits, and address potential risks and challenges.

Another question to ChatCPT: What should the European Union do to make sure artificial general intelligence is beneficial for humanity

The European Union can take a number of measures to ensure that artificial general intelligence (AGI) is beneficial for humanity, such as:

- *Developing and implementing ethical guidelines and constraints for the development and use of AGI*
- *Investing in research and development to support the advancement of AGI technology*
- *Encouraging collaboration and partnerships among researchers, scientists, and technologists working on AGI*
- *Promoting diversity and inclusion in the scientific and technological fields to ensure diverse perspectives and prevent bias*
- *Providing regular monitoring and evaluation to ensure that AGI is behaving in a safe and responsible manner*
- *Encouraging transparency and accountability in the development and use of AGI*
- *Ensuring that AGI is developed and tested in a secure and safe environment.”*