




Augmented Intelligence Summit

Introduction to Earth 2045



“In order to
know the world,
one must
construct it.”

– Cesare Pavese

Meet Earth 2045

The model world of 2045 being explored at AIS, denoted “Earth 2045” for brevity, is defined by a number of assumptions. These are not predictions, nor necessarily goals, but rather just defining characteristics of the plausible and aspirational but fictional world we are fleshing out and investigating at the Summit. Terms in bold italics are fleshed out below in the **Lexicon**.

Artificial General Intelligence is a reality, but has been carefully managed so as to avoid individual AI agents of far beyond human intellect. Rather, there are general-purpose systems of AI services under human control, as well as many individual AI systems and agents of varying capability, many comparable to humans in general reasoning capability; and more narrow AI is ubiquitous.

At the highest “top tier” level of capability, **Comprehensive AI Services** systems are used to maintain global political, technical and economic stability in three historically-generated spheres of influence: US, European, and Chinese. These service providers are human-controlled but capable of executive directives with dramatically trans-human capability. On the geopolitical level these human-AI systems function much like multi-nation mutual defense treaties, and enjoy a monopoly on nuclear weapons and other large-scale uses of force. They have negotiated a stable set of

verifiable agreements under which they interact while maintaining different “flavors” of AI services in their spheres of influence. These systems also act to prevent the development of rival **superintelligence** systems or agents, and enforce laws and regulations governing AI systems at lower capability levels.

It is generally believed -- and an explicit aim of major AI companies -- that extant AI systems are not sentient or conscious, and do not require moral consideration. AIs that appear to have feelings are generally frowned upon outside of particular contexts. It is illegal to deceptively impersonate humans through voice, text, or video. To the degree that computers respond on behalf of humans, it is mandatory to disclose that, with particular regulatory frameworks governing human-AI interactions in the realms of finances, health, and law.

The AI revolution has created a massive economic boost globally relative to 2019, with global economic growth (especially in the information sector) of ~10%/year on average, and by some measures up to 20% in the early 2040s. However several powerful mechanisms including the **Windfall Clause** and **Windfall Trust** have allowed a widespread sharing of this wealth so that global inequality is substantially lower than in 2019 even while GDP has grown tenfold.

In most jurisdictions people enjoy significant legal protections against some of the potential adverse effects of AI. these include:

1. In the EU, US, and many other jurisdictions, a comprehensive **Data Ownership Provision** has been enacted so that people have broad and comprehensive rights to the data they generate, and how it is shared and monetized.
2. To the extent that behavior modification caused by AI is grossly negative (e.g., addictive), it must be disclosed upfront - similar to the way alcohol and cigarette have warnings associated with them.
3. There is a strict set of legal and technical standards defining **AI Fiduciary Assistants** that are required to act on their client/owners behalf. These have become relatively ubiquitous and while they vary in power, even the very poor can afford or be provided an assistant that can act in their interest on many matters.

Key sectors of the economy have been largely transformed by AI services; for example:

1. **Transportation** is now largely automated, with intelligent mobility interoperability standards in place that allow (now nearly-universal) autonomous vehicles to communicate and coordinate.
2. **Medicine** has been revolutionized by a combination of AI-assisted medical research, advances in genomics, and machine learning applied to massive datasets that allow

Chinese “Forest City” that eats smog.

highly personalized and effective medical treatment.

Most of the globe is **de-carbonized** by 2035 thanks to improvements in monitoring technologies, international regulatory frameworks, and AI-assisted breakthroughs in materials science and nuclear (fission and more recently fusion!) reactor design.

Yet all is not rosy...

Job automation and displacement have caused massive social issues; there are significant tensions between the Chinese and US/EU blocks, with the latter considering the former to have fallen into a full surveillance state; environmental degradation, resource pressure, and species loss, while not catastrophic, has still been extreme. AI-assisted physical, social, cyber, and memetic attacks have caused significant disruption and loss of life worldwide.



“The most interesting question of all to me is not to sit and speculate about what is most likely to happen but rather to ask what we would like to happen. What sort future for life would we like? And then think about how to steer in that direction. I feel that today, media portrayals of the future, especially from Hollywood, are almost always dystopic. It’s ridiculous. That’s a really bad strategy for actually getting to a good future.”

- Max Tegmark

Lexicon

For the purposes of clarity of discussion at AIS we adopt the following meanings for a number of core terms. These are not universally-agreed-upon definitions but are proposed as a means to avoid spending time on semantic debates or on talking past one another.

Artificial intelligence: Historically, a moving-target definition of computing systems designed to perform the sorts of tasks that human cognition but not contemporary computers perform. More recently, machine systems characterized by assigning a goal/metric of what is to be computed, without an explicit algorithm for doing so, and that can potentially learn and adapt to new situations and problems.

Artificial narrow intelligence: AI systems with a particular/narrow set of capabilities. Often these take the form of a system trained on human-generated data (or in a game/environment) that then perform a task such as image/speech classification, navigation, etc., without significant further learning or flexibility. All circa-2019 systems are of this type.

Artificial general intelligence: AI systems that can -- like humans -- flexibly learn and reason over a variety of domains (though at varying levels of capability). AGI systems so-named at the AIS would for example be capable of passing a challenging Turing test (if so configured), learning a new computer game at human or better speed, performing well at

human standardized tests, and explaining at a human-comparable level how they arrive at results. They would also be capable of substituting for humans in an array of tasks and occupations.

Superintelligence: a term defining an AI system that is in its totality dramatically more capable than any individual or (AI) unaided group of humans and essentially all cognitive tasks. In Earth 2045 the “tier 1 comprehensive AI services” human-machine systems would constitute Superintelligences, each capable of out-thinking any circa-2019 cognitive system (including human society as a whole).

Tool AI and Agent AI: it is useful to distinguish between agent AIs, which like humans (or animals) have goals that they act in order to achieve, versus tool AIs that are explicitly constructed to not follow their internal goals, but to achieve the goals of a human. Most AIs in Earth 2045 are constructed or required to be tool AIs.

Sentience: for our purposes, having a subjective experience of potential positive or negative valence , requiring moral consideration.

Windfall clause: a hypothetical agreement signed by AI companies that above a certain (very high) level of annual profit they would donate funding to nonprofit/public charity use.



Windfall Trust: a hypothetical organization that would receive funds from signatories to the Windfall Clause, and use/distribute funding for the benefit of its shareholders, which would be each person on Earth, each holding one share. The Windfall Trust provides, among other things, a level of universal basic income.

(Comprehensive) AI services: AI systems characterized by high modularity in which each module is a tool that may be called by other modules or by its governing humans. There are multiple motivations for this including preventing a single “black box” intelligence that develops instrumental drives misaligned with human goals or wellbeing. These services are “comprehensive” if the system as a whole constitutes a high-level AGI or superintelligence.

Fiduciary AI assistant: A (hypothetical) personal AI assistant that is highly capable relative to current-day SIRI, Alexa, etc., and is both designed and legally required to act in the interest of its “owner,” like a lawyer or financial advisor and with comparable ethical overrides (so that for example you can ask this assistant to commit an obvious crime; it won’t do that but it will inform and dissuade you rather than report you.) This assistants would be both highly customized and capable of learning its user’s preferences, and work to protect and forward the user’s interests.

World building: The creative process by which an individual or group develop the setting of the world in which characters will interact. This includes infrastructure, economics, politics, education, transportation, social systems, etc. -- all aspects of a world that a character might interact with. This enables us to better understand the motivations of characters in this world, what the characters do, and how they will respond to events.

Data ownership provision: Individuals have rights to the entirety of their cyber footprint. Individuals decide who can use data about themselves or their online experiences. Parents own their children's data until they're 16. From 16-18, parents and children have joint ownership and both parties must agree to how the data is used. At 18, individuals gain full ownership of their data. It is illegal for companies and websites to say that a person must give up the rights to their data in order to visit or interact with the company or site.

Existential threat: Any event that could cause the deaths of 90% or more of life on earth, especially 90% of humans, or any event that dramatically reduces quality of life, such that all of society loses food, electricity, clean water, etc., for a time longer than humans can survive.

Existential hope: The idea that we can achieve a future society in which the beings have everything they could possibly need and want, and we've eliminated unnecessary suffering, or perhaps all suffering. This is a future ideal that is so grand it can encompass everyone's ideal.



Timeline

The following is an extremely incomplete timeline from present-day until 2045. One of the activities for AIS participants will be to discuss, debate, and fill additional color and detail into this timeline.

Year	Events
2021	US announced major AI legislation including funding for basic research, and soft regulation over multiple AI applications and use cases.
2022	Individual data ownership laws passed in EU
2023	First signatory of Windfall Clause
2024	<ul style="list-style-type: none"> Individual data ownership laws passed in US. Intelligent mobility interoperability standard released EU forms well-funded international center for the study of artificial intelligence (ICSAI)
2025	<ul style="list-style-type: none"> Windfall Clause signed by all major US, UK AI groups Windfall Trust spun up using corporate, philanthropic and govt seed funding
2026	<ul style="list-style-type: none"> EU regulations requiring identification/disambiguation of AI va. Humans China institutes massive government income redistribution program
2027	Mobility standards adopted by all major automotive companies and US govt. US regulations requiring identification/disambiguation of AI va. Humans
2028	In public event Google Deepmind agent passes a challenging Turing test
2029	Fiduciary AI assistant introduced as marketing concept by major AI company
2030	<ul style="list-style-type: none"> Google record \$1 trillion in annual revenue. ICSAI demonstrates a high-level AI system
2031	<ul style="list-style-type: none"> EU rules governing AI promotion of addictive/manipulated behavior; US follows one year later. Fiduciary AI standards developed by industry groups
2032	Fiduciary AI standards adopted by EU, US; such assistants now near-universal
2033	First triggering of Windfall clause, by a major US company;
2034	<ul style="list-style-type: none"> Net global carbon emissions reduced by 75% relative to 2019 Windfall Trust at 1% of GDP.
2035	NATO treaty is expanded to include provisions governing high-level AI systems.
2036	
2037	US government contracts with Google to form national Strategic AI center (NSAI)
2038	
2039	Windfall trust at 5% of global GDP
2040	
2041	US's NSAI, ICSAI, and Chinese government announce a threefold AI-mediated treaty.
2042	
2043	
2044	Windfall trust at 20% of global GDP