Artificial Intelligence : What in the world are we playing at?

A summary of a talk by Robert Whitfield at the U3A monthly meeting of 12 February 2026

Robert has been involved in the world of AI for the last 10 to 15 years. His involvement has been varied and through such organisations such as the Transnational Working Group on AI, The World Federalist Group, the Institute for Globalist Policy and GAIGA*Now*

Introduction

We are in exceptional times and so I do not want to only focus on the problems and issue of AI, but and what we are doing about it. And in the context of the world that we live in today, not a few years ago.

I have admired the work of Gerd Leonhard who talks about a good future but also the problems we need to overcome – The Bad Future (Power over Purpose, Technology over Humanity and Transactions over Trust). The good future will not be made in America. The volatility of the Trump administration has lead us to a position where truth becomes irrelevant and might equals right. International organisations are ignored.
We are in a very extraordinary and disturbing period in human history at the moment.
And in that human history, we have refined ourselves with artificial intelligence.

With that background the talk covered :
What is Artificial Intelligence?, Catastrophic and existential risks,
Taxonomy of AI Safety Risks, Dangerous behaviour ALREADY observed,
How fast is AI moving?, When is AGI expected? Perceived scale of advanced AI risk,
Other AI issues, Solutions to safety challenges, Launching GAIGA*Now,*
Enforcement of International Treaties and Conclusion

What is AI?

There are many different definitions of what artificial intelligence really means, but the most useful one is the statement that it is the study of systems that perceive their environment and act to maximise goal achievement. This is from the book by Russell and Norvig, that has been used of 180 countries to learn about artificial
Basically, the development of the power of artificial intelligence that you see is not from programming, it is from making it more powerful than the people who are developing it. Developers have discovered what increased power can deliver and they are often surprised by the results. It is not because they've actively programmed something to do something. They have played around, tweak this, double that increased that, and they find the resulting capabilities. So, they are grown, not crafted.
At some time stage, therefore, we should expect the machines take control.
So said Alan Turing in the 1956
And the question that our generation have to decide is whether we wish machines to take control, and if not, what are we going to do about it? So that leads on to reflecting the idea of major risk.

Catastrophic and existential risks

Up until a few decades ago our biggest threats were all natural in terms of asteroids or super volcanos or whatever. But we have begun creating existential risks of our own The first people were aware of that was the nuclear bomb. But we are currently in the process of developing others, and it is argue by many that the greatest threat at the moment is that of uncontrolled advanced AI.

Taxonomy of AI Safety Risks

What we should be most concerned about is the pace of change. Comparing the ChatGPT capability that we experience today with what it was a few years ago, If it continues to develop at that rate what can it become? And that is what we have to be concerned about. And we could not simply wait till that happens and then try and do something about it because our options may no longer exist. We need to think about doing something about it NOW.
Examples of the major risks include :
Humans / bad actors, or machines, Military or Civilian,
Losing control of advanced AI , Advanced AI in the hands of bad actors, gradually slipping into an AI-dominated world, Malicious use
AI race, Organizational risks, Rogue AIs, Misuse, Misalignment, Mistakes, Structural risk

Dangerous Behaviour Already Observed
Scheming: models are introducing subtle mistakes into their responses,
attempt to disable their oversight mechanisms ,surreptitiously copy what they believe to be their model weights to external servers.
Also, Large Language models can be trained to outwardly appear aligned with new directives request while internally preserving their
own, potentially harmful preferences.
They can be insider threats, e.g., blackmailing the engineer in
charge of shutting them down.

In effect starting to take control away from developers

How fast is AI moving?
Dario Amodei, CEO of Anthropic:“The single most important thing to understand about AI is how fast it is moving.”
Machine intelligence capabilities have been assessed at growing from 1952 to 2018, at roughly 30% p.a., and since 2018, at around 300% p.a. This pace is increasing.
In comparison the difference in the number of neurons between a chimpanzee and a human is about 3 and it took evolution 5 to 7 million years to create that difference.

When is AGI expected?
AGI (Artificial General Intelligence)is a theoretical advanced form of AI that can understand, learn, and apply knowledge across a wide range of tasks at a level equal to, or exceeding, human intelligence,

The timing of the arrival of AGI) has been forecast
by several leading AI Executives as varying from 2026 to 2030.

Perceived scale of advanced AI risk

Toby Ord has estimated that general artificial intelligence presenting a chance of 1 in 10 of humanity no longer existing within the next 100 years. He certainly considers that that percentage risk has increased in the last 6 years. Jeffrey Inchon talks about a 20% extension risk within 10 years.

Other AI issues

Ethics**:** AI currently poses serious ethical risks, including
bias, threats to privacy ,surveillance, misinformation

Equity: The development of AI is currently controlled by a small number of companies and states, exacerbating the current global power imbalance and hindering a more equitable balance of resources.

Interoperability: AI is a cross-border technology, where differences in regulation greatly complicate the determination of liability and effective interoperability.

Together these issues emphasise that any solution must be international

Solutions to safety challenges

Solutions being pursued are both Technical and Governance
and preferably a combination of the two.
The risks are currently growing far faster than the capability to contain them.

Governance solutions such as:
Voluntary agreement between the major AI companies, backed by States,
Global AI Framework Agreement

Technical approaches such as
FLOP based thresholds, Hardware – modified AI chips enabling privacy-preserving verification and enforcement mechanisms between states.
Others include Alignment (Russell et al), Scientist AI (Bengio),Creating an Open Agency AGI (Drexler)
You can see a short report summarizing 14 “solutions” in the International AI Governance solutions report on www.gaiganow.org

Launching GAIGA*Now*

Several international AI summits have been held such as:
Bletchley Park AI Safety Summit – November 2023, Seoul AI Summit – May 2024,

Paris AI Action Summit – February 2025, Next AI Summits: India February 2026, Switzerland 2027.

The AI Safety Movement is not winning – and this needs to change. The voices for international AI governance are too diffused and need to be brought together.

One avenue for this is the launch of the Global AI Governance Alliance (**www.gaiganow.org**). Established in 2025 with the aim of A Global AI Treaty, possibly preceded by a US / China agreement on AI Safety

Enforcement of International Treaties

The issue of enforcement is likely to be critical for the future of humanity
There is no universal enforcement body at present

Treaty enforcement relies on a variety of methods, including:
National Courts, International Courts and Tribunals, Treaty Bodies, Diplomatic Pressure, Sanctions and Public Opinion.

Specific provisions for are required monitoring compliance and addressing potential breaches, e.g.: Reporting Requirements, Complaints Procedures, and Inquiry Procedures.

Treaty enforcement faces numerous challenges, including:
Political Will, Complexity and Sovereignty

How do you enforce a clause in a Treaty that a state can simply walk away from?

Given the prospect of advanced AI putting the very existence of humanity at risk in the coming year, is a Federal World Government the necessary tool to provide safety?
Baruch Plan for Nuclear energy proposed it in 1946, but it was not fully thought through, and not accepted

A Baruch Plan for AI – would need to be fully thought through and embrace central military control. But how do we avoid the downside risks of such a solution?

Conclusion Global

Advanced AI offers huge capability – and currently threatens the future of humanity.

All those supportive of international AI Governance need to work together to bring about suitable international AI Governance covering:
An early US / China agreement re advanced AI and a global AI Treaty

The Global AI Governance Alliance seeks to contribute to include tribute to these crucial steps.

The Enforcement of a resulting global AI Treaty will be key to our future

Conclusion Closer to Home: YOU

What can YOU do?
The future of humanity rests on what the citizens of the world do in the coming years and decades
Take what action you can, learn more about these issues and apply your own mind, contact your MP and express your concern – use CONTROL AI's access to MPs tool as a default, share your views with friends and colleagues, engage with a relevant civil society organization, write a blog or an article if you feel you have something to say.

Action leads to a sense of empowerment, it has a positive impact on the issue itself and reduces anxiety.