

Can AI Save Humanity? – Maybe and here's how

Colin Irwin

In the age of thermo nuclear warfare and the prospect of MAD – [Mutual Assured Destruction](#) - the father of modern day sociobiology [E. O. Wilson](#) tells us that: “The real problem of humanity is the following: We have Palaeolithic emotions, medieval institutions and godlike technology. And it is terrifically dangerous, and it is now approaching a point of crisis overall.” That danger may now be even more acute than E. O. Wilson imagined with the invention and application of Artificial Intelligence to the management of human affairs. Some fear AI or Artificial General or Super Intelligence could herald the [demise of humanity](#) while others suggest it might be [our saviour](#). I take the view that it could be our saviour if we choose to make it so - [here's how](#).

In the 1990's I developed '[peace polls](#)' to bring the views of 'the people' into the negotiation of the Belfast / Good Friday agreement. It worked very well and helped the parties in Northern Ireland to achieve an historic peace that has lasted to this day. The same methods have been applied around the world since then, with mixed results. The most notable success was [Macedonia](#) where the people wanted new elections that kept Macedonia out of the Balkan wars. While the most significant failure was [Israel and Palestine](#) where the peace polls detailed an acceptable peace agreement, for both Palestinians and Israelis, but Prime Minister Benjamin Netanyahu turning it down resulting in more war. The problem is this, human conflict is not monitored everywhere at all times and we are not told what 'the people' can accept as a peace agreement, as a normative part of all diplomatic discourse. Can AI do this when the human political elites refuse to cooperate as effective partners for peace?

Recent AI developments suggest this is now possible. In June 2020 the [United Nations Innovation Cell](#) were ready to test a new AI platform, [Remesh](#), that could engage up to a thousand participants in a digital dialogue and compute what they could agree to, in real time. I was brought in as a consultant and we got very good results in [Yemen](#) where traditional peace polls got essentially the same results, but they took weeks to run. Our AI results were instantaneous. We then moved on to [Libya and Bolivia](#) with equally good results. Significantly, [in Libya](#), we published what the Libyan people could agree to on Facebook and TV reaching a 1.7 million audience in real time, leading to the formation of a Government of National Unity (GNU). Clearly this should be done everywhere there is conflict but the UN, sensitive to criticism from Member States and a possible veto from the Security Council hardly ever make the results of these peace dialogues public and do not run them where they might not be welcome by a Member State. Such as Israel and Palestine. Perhaps Artificial General Intelligence (AGI) is the answer?

In May 2023 [OpenAI launched a programme of research to make the governance of AI more democratic](#) or, in the terminology used in the industry 'alignment with the will of humanity'. The problem of writing a constitution for AI governance, that most people could agree to, was no different to writing a peace agreement between conflicting parties. From an AI

perspective the process is the same. So with [Andrew Konya](#), the CEO of Remesh AI, [Lisa Schirch](#) of Notre Dame University, [Aviv Ovadya](#), and myself we wrote draft [AI governance proposals for medical, vaccination and war and conflict issues](#). With real time input from the public and relevant experts we could do this in about two weeks for AI governance. But with AGI the process could be fully automated with an acceptable peace agreement written in days or hours. The other major AI company, Google with DeepMind, have started to do this with their [Habermas Machine](#) and no doubt the new Chinese AI start-up DeepSeek will also be able to do this. But [DeepSeek will not answer questions about Tiananmen Square](#). It comes down to this. Who will control AGI?

On the 20th of September 2024 the [UN published their recommendations for the future governance of AI](#). And most significantly they said: “AI governance should be anchored in the Charter of the United Nations, international human rights law and other agreed international commitments such as the Sustainable Development Goals (SDGs)”. This should work OK. With the UN Charter as the ‘objective’ of AI governance and international law the ‘rules’ to meet those ‘objectives’ [AI can be aligned](#) to do nothing but write agreements and make recommendations to keep the peace, everywhere, all the time. Providing this ‘godlike technology’, to use E. O. Wilson’s phrase, is allowed to do so without a [UN P5 Security Council veto](#). We have a choice to make, as do the owners and managers of the AI industry.