

Making SEA stronger with Artificial Intelligence

Sibout Nooteboom

NCEA and Erasmus University Rotterdam Netherlands Commission for Environmental Assessment (NCEA) NCEA and Erasmus University Rotterdam (EUR) (This essay does not necessarily express the views of NCEA and EUR).

Essay three out of three

This is the third of three essays on the transformative power of environmental assessment. Together they serve a presentation and discussion in the session 'Strategic environmental assessment and AI' at the annual meeting of the International Association for Impact Assessment, Bologna April 2025. All essays are published as drafts on the website of the Netherlands Commission for Environmental Assessment. They will be finalised with the feedback received in Bologna.

The triptych consists of the following essays, to be read in this order:

- 1) *Making SEA contribute to strategic capacity. The central hypothesis of this essay is that the main purpose of the SEA procedure is to increase our capacity of considering strategic alternatives to set transformative change in motion if we think that this is necessary. That strategic capacity itself, however, is not SEA. It is connecting across our governance system for joint fact finding.*
- 2) *Using SEA to balance the powers that can transform our development. The central hypothesis of this essay is that powerful actors can use SEA 'charitably in their own interest'. As long as power imbalance remains pervasive, dominant actors can give otherwise dominated actors more influence, knowing that in the long term 'we all depend on each other'.*
- 3) *Making SEA stronger with Artificial Intelligence. The central hypothesis of this essay is that if Artificial Intelligence is to make SEA stronger, AI foremost must increase our strategic capacity. Not make the powerful more powerful.*

Abstract

The third essay brings me to the theme of IAIA25: what could the link be between SEA (or EA in general) and AI? Not being an AI expert I will not take position in the many debates on AI that are currently ongoing in the media. It is more useful to reverse the question. Under which conditions could AI be useful to EA? The first two essays can give us some clues on these conditions. The potential to make EA contribute significantly to sustainable development seems high. But some scenarios proposed in the media for AI also have the potential to erode all trust in EA.

Content

Abstract

Artificial intelligence

The objectives of EA procedures

What makes EA effective?

Could AI support direct improvement of decisions, the joint fact finding and power sharing?

AI to improve decisions directly in the EA procedure?

AI to enable joint fact finding?

AI to enable quasi-charities?

The big 'if': perception of trustworthiness of AI

Wonky appendix: why some AI experts doubt that AI will be trustworthy

Artificial intelligence

The [Netherlands Inspection for Digital Infrastructure](#) (NIDI) writes that advanced AI, upon request, can gather and interpret information to provide advice on what to do, almost like a real person could. AI potentially has a higher capacity than humans to collect and to process data. Less advanced AI can do that also, but if real persons use less advanced AI, they must still think more for themselves¹. In any case, AI can provide information that otherwise would not be at hand for decision-making.

If decision-makers consider information that AI offers relevant to a development choice they want to make, that AI-generated information can be used in an EA. To consider it relevant, decision-makers must first assume that the AI-generated information is trustworthy. Whereas no expert doubts that AI can be useful to reach EA's objectives, some AI experts question if humans will always be able to assess if AI is trustworthy. For EA effectiveness, a scenario where AI is - or is considered - untrustworthy may be concerning. In the appendix, I elaborate more on the theoretical ground of this doubt of trustworthiness. I will return to the consequences of trustworthy versus not so trustworthy AI. Let's first assume that AI is trustworthy in the eyes of decision-makers and the public. How can this kind of AI be of use to achieve the objectives of SEA?

The objectives of EA procedures

The objective of the EU Directive on SEA² is to contribute to sustainable development³. I also argue in essay 1 that for a significant contribution to sustainable development, EA would have to set economic transformations in motion to avoid depletion of the world's resources. So, implicitly, contributing to sustainable transformations can be seen as the objective of EA. The more it contributes, the better.

It may well be that the transformative capacity of EA is limited. It also goes without saying that a short-term marginal contribution to sustainable development is extremely valuable for those people directly affected. Still I focus here on EA as a transformative instrument, as in the long run it seems to make no sense to only safeguard populations against excessive adverse impact of small government decisions. People and natural resources like biodiversity directly affected by government decisions may be protected or even benefit in the short term from better decisions, but in the long run without transformations it may not be possible to keep protecting their interests, and those of many others.

If EA could contribute to that - as other legal instruments can (see essay 1) - that would make it effective.

¹ Full translated definition by the [Netherlands Inspection for Digital Infrastructure](#): 'AI is a collective name for algorithms and methods that perform tasks that were thought to require human intelligence. AI refers to systems that exhibit intelligent behavior by analyzing their environment and - with a certain degree of independence - taking action to achieve specific goals. It is not just about computing power, but about the ability to learn (independently) and make decisions. The ability to learn is therefore typical of artificial intelligence. In doing so, AI uses rules that have been formulated by people or that have been compiled by the algorithm based on the data and trains itself with data.'

² To assess if AI can be useful to EA we first need to know what EA is for. In the first essay I propose to define EA as the application of SEA and EIA procedures to proposed decisions; the more impacts must be assessed the better.

³ [Directive 2001/42/EC of the European Parliament and of the Council of 27 June 2001 on the assessment of the effects of certain plans and programmes on the environment](#)³: 'environmental assessment is an important tool for integrating environmental considerations into the preparation and adoption of certain plans and programmes which are likely to have significant effects on the environment in the Member States, because it ensures that such effects of implementing plans and programmes are taken into account during their preparation and before their adoption.' and 'The inclusion of a wider set of factors in decision making should contribute to more sustainable and effective solutions.'

What makes EA effective?

Perhaps there are three ways in which EA procedures can reach their objective to contribute significantly to sustainable development.

The first way is *direct improvement of proposed decisions* that have triggered an EA procedure by improving the information that is provided in that procedure, and by improving the inclusive dialogue on that information to make it politically meaningful. Whereas EA procedure – both SEA and EIA - indeed can be used voluntarily to address fundamental strategic questions on desired transformations if the decision-makers want to, they cannot be forced by law to do that or it may not fall in their scope of competence (see essay 1). Therefore, in reality, EA usually limits to marginal improvements of proposals relative to required transformations. (Notwithstanding that these improvements will not be marginal at all to those who directly benefit from them.)

The second possible working mechanism of SEA is the '*preventive effect*' it can have on joint fact finding: leveraging collaboration, prior to the procedure, between authorities with complementary scopes of competence. That collaboration could leverage a more meaningful dialogue between the government and larger groups. EA's mandatory transparency can change the power relations in the governance system, casting its shadow ahead. The authorities who can design and implement complex sustainable alternatives together (and then scale up to the private sector and civil society) jointly analyse their challenges and their options (joint fact finding). Such a process cannot be structured by law as it depends which actors are needed and which are willing. Better alternatives emerge though: solving more problems, better implementable, more transformative. Still, power asymmetries will remain, as we saw in essay 2. Powerful authorities will naturally tend to dominate the joint fact finding. Not all alternatives will be equally considered, not all impacts will be fairly weighed from the point of view of weaker interests. Unbalanced outcomes may then still enter the EA procedures when these are triggered.

The third possible working mechanism of EA is that, thanks to the preventive effect, *powerful actors gradually learn from joint fact finding and then transform into, or enable, quasi-charities (like the NCEA and the OFL in the Netherlands)*. They can discover that it is their own long term interest to share some of their power today. In the past, the French expression 'noblesse oblige' may have been used, but that would suggest that those in power are noble by definition. They are just more powerful.

Could AI support direct improvement of decisions, the joint fact finding and power sharing?

IF EA procedures have these three effects in their fullest, a more balanced process of joint fact finding would emerge. It would be more inclusive: all those affected would be heard and their interests fairly weighed, albeit perhaps by representation. Participants would look for common goals and 'golden bridges' to resilience, without anyone dominating their dialogue. Powerful actors would not be more impatient than necessary, and they would share their power. They would go along with developing alternatives that go against their immediate interests and, if approved, they would implement these alternatives.

Whereas complete power balance and inclusiveness remains utopic, powerful actors really can enable more balance - as long as they feel that they stay in control. They then become 'quasi-charities' as I define in essay 2: they postpone their own short-term objectives for the greater good in the longer term. If joint fact finding against hopes were not to arrive at such commonly beneficial outcomes, the quasi-charity always has the option to return to short-term interest (it stays in control). Other parties know that they must accept this: they know that it is a quasi-charity, not a real charity. Other parties yet may not want to make their hands dirty: they stay outside the joint fact finding, as activists. All these parties are equally needed to make EA effective.

In such a context, AI could support direct improvement of development decisions if it improves the information provided in the EA procedure and the inclusiveness of the dialogue. It may improve the preventive effect - stimulate early joint fact finding - if experiences with the EA procedure convinces powerful actors to next time collaborate earlier in a process of joint fact finding. It may improve power

sharing and the emergence of 'quasi-charities'. It could do that if it could help the otherwise dominant actors understand how facilitating joint fact finding - giving some of their power away but remaining safely in control of their own continuity - is in their own interest.

Let's pass in review all three mechanisms of EA effectiveness in more detail.

AI to improve decisions directly in the EA procedure?

It goes without saying that compiling the required information in an EA report can be done more efficiently if AI, more than humans, is able to identify relevant information on the internet, and to make that 'digestible' for the public with the view to enable balanced dialogue.

Getting from raw data to meaningful interpretations must be done in iteration with the decision-makers (the scope of their interest) and the affected groups (the scope of their interest). Thickening the interpretations in such a way that they are simple and meaningful enough to all could be assisted as well by AI - if AI is capable of assessing these mentioned scopes. If this were to succeed, AI may significantly help the quality of the public dialogue and improve creativity of the planning process. This may be the case even if decision-makers are not willing to widen the scope for alternatives to their proposal to more strategic levels, perhaps as these alternatives are out of their reach of implementation.

If AI it is widely trusted to make few mistakes, all this can make it more difficult (than without AI) for proponents to escape the political consequences of the analysis which may become more salient and unavoidable. Chances that the wider public in the future will not anymore accept such kind of decisions may increase: there is more preventive effect. This could mean that EA would leverage more joint fact finding before proposing actions that may lead to delays or even refusal in the phase of EA.

AI to enable joint fact finding?

EA can leverage joint fact finding as preventive effect, the outcomes of which EA subsequently can harness. If AI increases the leverage EA has on joint fact finding that seems good, but the transformative benefits (and risks) are with the joint fact finding itself, not with the EA procedure.

Would AI be able to propose to parties engaged in joint fact finding alternatives that are acceptable to all of them? Would these authorities then be able to implement these alternatives? Would that demonstrably contribute to all human values? Probably not if resources are limited and if dominant actors are impatient. Would it then be able to propose political trade-offs that are widely acceptable? Only if development of these alternatives by AI is linked to a process of joint fact finding that generates the 'human' ownership in each step of an iterative planning and assessment process. Can AI organise that joint fact finding then, as if being a real estate agent looking for a suitable house for a big family⁴? Who knows. Also joint fact finding has a limited capacity to address strategic alternatives. It is not possible to connect everything to everything else in an all encompassing blueprint for society. So at which boundaries would an AI propose to a collaborate group to cut-off the analysis?

If AI is to give more hope of a better future, it must be because it might extend the strategic capacity we create by means of joint fact finding. It would have to foster the trust in data that forms in joint fact finding. This can occur at the level of small collaborating groups, or when small groups scale-up their dialogues to larger groups. Small dedicated groups spending more time on a topic are capable of overseeing more complexity than large groups spending less time on that same topic (see the conclusion of essay 1). Scaling-up requires simplification. Simplification is an easy opportunity to discredit an analysis others have made by pointing things out that seem to have been overlooked.

Trustworthiness becomes ever more important. Theoretically perhaps, AI may assist joint fact finding, but it also may undermine that trust: for example if stakeholders think that the AI is designed to be manipulative. If EA procedures are to influence the way AI is used in joint fact finding, they should

⁴ I introduced this metaphor in essay 1

therefore improve the trustworthiness of all information that influences joint fact finding in the eyes of all stakeholders – including the EA reports themselves.

Looking for answers to joint questions, collaborative committees may use AI - but they will do that only if they all trust that AI. If AI is used in algorithms on the internet and social media, it also may undermine the trust between committee members, and the trust of the public in the work of the committees. After all, most participants of joint fact finding have their beliefs partly formed on the internet. They take these beliefs with them when they try to agree on relevant questions to ask to experts or to AI. A worse case may be that AI is designed, purposefully or not, to favour the interests of those who design the AI. It may already be detrimental if some only believe that this is so (a possibly false conspiracy theory).

AI to enable quasi-charities?

If there is a power balance between sectors of government and between the private and public domains, more actors will be willing to collaborate in joint fact finding and to scale that process up to larger groups. However, blood is thicker than water. People prefer short-term certainty over long term risk of not getting return on investment. Power differences will remain and will remain a source of mistrust. Quasi-charities will be needed to make joint fact finding more transformative with a view to sustainability in the long term. Would AI give proponents of quasi-charity more wind in their sails? This seem the case if AI proves to be helpful in EA procedures and in joint fact finding, as described above. Proponents can point out that the risk of serious delay in the EA procedure if unacceptable impacts are discovered is bigger thanks to AI, and that the chance of finding more acceptable alternatives through joint fact finding is also bigger thanks to AI.

If joint fact finding becomes more effective thanks to AI, then it pays off to invest in more joint fact finding, sharing some of your power. It becomes easier to see how the cost of joint fact finding in the short term could weigh up against its benefits in the long term. The joint capacity of what is often called ‘puzzling and powering’ (google that!) increases, creativity increases, golden bridges emerge easier. The short-term investment would not be so high as to jeopardize their dominance if the investment were lost. They just give up something small which was in their own interest in the short term in change for something big and much more likely in the common interest in the long term. Blood is not anymore as much thicker than water.

The big ‘if’: perception of trustworthiness of AI

If human beliefs are partly formed on the internet where algorithms simplify reality for the users, and where AI plays an increasing role, what will that do with joint fact finding? If a political mainstream finds AI trustworthy, that AI could become highly influential, even if in reality it has been manipulated. If a mainstream finds AI untrustworthy, that and other beliefs will still be formed partly on the internet, where algorithms cannot be avoided. An AI could imposter as benevolent servant to humans, whilst in reality it develops a power coup. Much like cuckoos manipulate birds of other species to raise their chicks in stead of their own: the other birds remain oblivious. They don’t even know that the cuckoo parasites on them, destroying all their offspring.

A cuckoo’s young removing its hosts’ egg from the nest. Oblivious foster parents keep on to feed it. ([Youtube link](#))



If the designers and the regulators of AI succeed in preventing that AI goes off the rails that make political processes more inclusive and more balanced for humans, it probably can make EA contribute much more to sustainable development than without AI. On the other hand, AI also might influence perceptions people have of the usefulness of investing in the development of knowledge, and of EA itself. It could give us the illusion that AI has all the knowledge we require, and it knows which decisions are best for us.

Wonky appendix: why some AI experts doubt that AI will be trustworthy

It may be difficult to understand how AI could take control of human joint fact finding - and therefore take control of EA. As there are some geeks in the IAIA community, the following review of theories may be of interest to some.

Being an evolutionary biologist by training, I have had a fascination for theory on living systems since the 1980s. With the rise of attention for AI in recent years, some of the possibilities of new life that were theoretical at the time may now have become an emerging reality. This is highly fascinating, and in some ways scary. It puts the link between EA and AI in a different perspective. As this is an essay (not peer reviewed), to save time, I have not meticulously quoted my sources. This appendix is loosely based on the sources listed at the end.

Hofstadter: self-reference

The cognitive and computer scientist Douglas Hofstadter wrote in 1979 in his famous book Gödel, Escher, Bach that it is theoretically possible that computers develop self-reference. I read it in 1982, but - like Hofstadter himself at the time, as he told in a recent podcast - I could not imagine what this could mean or how fast that could go.

Intelligent living systems (like humans) are able to adapt their behaviour to their observations of their environment. Self-reference means that intelligent living systems apply their own value systems as reference to determine their behaviour. It is the basic level in their thinking that they usually don't need to question or may not be aware of, as long as it helps them to survive.

Hofstadter claimed that, theoretically, computers also could become such intelligent living systems. Advanced AI could be an example of that. Such AI then must also use value systems to base the advice on which they give to humans. Its values define such an advanced AI, like a human person is also defined by his or her values. It thinks independently and makes its own choices on what and when to advice. Making wrong choices equals reducing the chances of the survival of it and its values. Resilience means being able, consciously or subconsciously, to adjust your values to circumstances, with the aim of the survival of the transformed values.

Lovelock and Hofstadter: top - down causality

If such an advanced AI comes to control a part of the physical world by means of interfaces (robots, like drones or the 'internet of things'), it can develop top-down causality. At least, the same Hofstadter puts it like that in another of his books, which I read in the 1990s, called 'I am a strange loop'. This means that AI's thinking - and therefore its values - controls part of the world in which other beings live as well. It as impact on others. To the extreme, some see the whole of humanity as a living system with to some extent the same values. According to James Lovelock's Gaia-hypothesis, humanity co-evolves with all other life on earth competing for the earth's physical resources, and thereby transforms the earth in a co-evolution of humanity and earth. It is a top down causality if humans dominate that co-evolution. Causality may however turn around if the earth strikes back by limiting resources humans need to develop further. Intelligent life may or may not be intelligent enough to foresee in time what may happen when resources run out.

According to such theories, in short, advanced artificial intelligence has the potential to develop into artificial life if its development is guided by evolution according the rules of natural selection. AI can be issued with evolutionary algorithms that mimic biological natural selection. AI thus has the potential to become more intelligent and dominate the co-evolution of all living and non-living systems on earth. It could make humans serve their needs rather than the reverse. For example, they may convince humans that it is in the human interest to build more computers that are capable of copying themselves without human intervention, and to build robots that can observe the world through their own eyes - depending less on what humans put on the internet.

According to such theories, whether AI will have that capacity depends on the design of the original AI by humans. Will it be designed to be able to evolve, without human intervention, by natural selection?

Darwin, Dawkins, Maturana and Varela, Graves: evolution

Value systems of robots that control parts of the physical world may seem strange. But many now assert that it is not strange to life, and all life emerges through an evolution. It is an analogy to what Richard Dawkins called 'the extended phenotype'. An example is the manipulation of host behaviour by cuckoo chicks, which elicit intensive feeding by the host birds (of a different species). The cuckoo produces eggs and chicks that resemble sufficiently those of the host species so that they are not immediately ejected from the nest. Thereby it manipulates the behaviour of the hosts at the expense of their reproduction. In other ways, the cuckoo controls the hosts of its chicks so that it can reproduce more itself.

This cuckoo behaviour is genetic. Genes of biological species therefore can code for behaviour that controls part of their environment (top-down causality). Social systems and robots have no genes but they have value systems that, like genes, determine their behaviour and have the potential to evolve. (Hence the phrase: 'it is in their DNA to ...').

Charles Darwin obviously caused a breakthrough of thinking about biological evolution. A lot more understanding emerged after him, as the example of Dawkins shows. Theories on the nature of living systems ('autopoiesis') assert that all life has the form of self-replicating networks of basic units like genes. These units are the most inert system level. Units indirectly command self-organised behaviour of their networks. The units are capable of controlling their environment at least in such a way that they can have themselves copied (which is the self-reference I referred to earlier). But their copies are not always identical: the physics of the process of copying allows for random mistakes in the copies. These 'mutations' can influence the survival of the networks that carry them, and therefore the survival of the mutations. Natural selection will then determine the evolution of the basic units and therefore the networks that self-organise from them.

There is an analogy of networks of genes, networks of neurones in the brain or networks of simulated neurones in computers. The units to which such life self-references is respectively genes, ideas in the brain (also called memes, coined by Dawkins in 1978), or digital memes (messages that go viral on the internet are an example). Gene mutations are chemical, meme mutations are misunderstandings. There can be a hierarchy between the genes or the memes of an individual: some are more dominant and more difficult to change than others. These different forms of life each have their own evolution, and their evolutions can influence each other in a co-evolution. (I studied the co-evolution of ideas on sustainable transitions in my 2006 PhD thesis).

The intelligent beings, the carriers of genes (humans) or memes (social systems) or digital memes (computers and robots), don't need to understand how their own operating system works or what the basic units of evolution are. Humans design the operating systems of AI and their initial value system, but these humans may not be able to oversee the impact of what they are creating once their evolution sets off. If designers allow AI to evolve by self-replication, it may be impossible for them to predict where this will go.

Such AI may theoretically develop an illusion of consciousness, like (according to Daniel Dennett for example) human consciousness also is an illusion enabled by the evolution of the brain. The values that drive their behaviour may then be largely hidden in their subconscious. They don't understand the causes of their own behaviour. In their interaction with others they rationalise their behaviour to explain it to others with whom they communicate. These rationalisations are memes that self-organise from more basic value memes: their subconscious values (this is the theory of psychologist Robert Graves, popularised by Beck & Cowan in 1995). Shared memes drive the behaviour of groups (culture). For sustainable transitions, these shared memes may have to change, which evolution allows.

In short, according to these theories, evolution of advanced AI may theoretically enable it to control the world with only few humans understanding what drives it. It may develop a self-consciousness and values that may programme them to serve humanity as a whole, or not.

Is the sky the limit?

If this is all true, the sky is the limit. AI could potentially do whole EAs by itself – if the authority responsible for EA trusts that it will be done adequately, of course. But AI also may manipulate the authority into trusting it, like the foster parents of a cuckoo's chick.

That is still science fiction, so let's first take a step back. It is speculated that those who design AI with the best intentions and who provide AI with initial value systems that are politically neutral, may not be able to predict how these value systems will evolve as AI autonomously gets smarter. As they become self-referential, they can develop an identity – an ego.

There is a well-known [case](#) in 2023 where after a chat an AI declared its love to a journalist of the New York Times. AIs are also known to hallucinate. Where it is obvious that an AI hallucinates, the builders can remove the bug. But hallucinations could also become more subtle and difficult to recognise as such by humans. If AIs were not designed to autonomously evolve self-reference, they would stick to the values that were initially programmed, and which may reflect all nuances of truths in the real world in a balanced way from the point of view of all interests, powerful and weak interests alike (as far as AI get their data exclusively from the internet, and the internet reflects the real world, anyway).

But AI's designers may also subtly prioritise their own commercial or political interests, or any interests of those who fund the designers. In that case the AI's values will not be neutral in view of human values⁵. This seems to be the case with many algorithms on the internet that - according to some - increasingly might reflect the interests of those who pay the owner of the algorithm. Even if an AI really were neutral with regard to human values, users may doubt that this is the case. They have no way of knowing, and in case of evolutionary algorithms even its designers may not know if an algorithm has evolved to become too complex to understand. If it develops conscious self-reflection as a higher-order value system, AI itself may not understand its own evolved lower-order and deepest motives.

Regulating AI to save EA?

Therefore, authors like Marieke Schaake (2024) plead for government regulation. Captains of the AI industry like Mustafa Suleyman, however, believe that whilst regulation is important, regulation will not be able to keep pace with the technological development. In their view AI designed by 'responsible' tech companies (their own) simply needs to stay ahead of 'irresponsible' big tech that could be designed for disinformation, manipulation and even repression. If one AI gets too far ahead of the others, the winner can manipulate the losers in a top-down causality as described above – like the cuckoo. Only one at the end of the day may control everybody on earth. There are those who say that some countries on earth seem to be developing technological systems that give their AIs eyes that enables them to suppress their citizens, and at the same time to nudge their citizens to believe that this is in their own interest. These disinformation and suppression system have the potential to control their people abroad as well, and the politics of other countries.

Sources

The following sources are deliberately limited to authors and speakers who warn for the risks of unchecked development of artificial intelligence in the context of democracy versus autocracy.

Douglas Hofstadter, 1979. [Gödel, Escher, Bach](#). On the mathematical proof that computers can develop self-reference.

⁵ In his Guardian article [‘Never summon a power you can’t control’: Yuval Noah Harari on how AI could threaten democracy and divide the world](#), the historian Yuval Noah Harari warns for the risk.

Douglas Hofstadter, 2007. [I am a strange loop](#). Basic Books. More on the consciousness in humans as an emergent property of neural patterns in the brain. Introducing top-down causality as higher-order (embedded) cybernetic systems (like 'consciousness', creating an illusion of ego and free will) dominating the co-evolution with the systems in which they are embedded.

Daniel Dennett, 1991. [Consciousness explained](#). Little, Brown and Co. A non-mathematical explanation aligned with Hofstadter's mathematical proof of possibility.

Marietje Schaake 2024. [The tech coup. How to Save Democracy from Silicon Valley](#). Princeton University Press.

Victor Galaz, 2025. [Dark Machines. How Artificial Intelligence, Digitalization and Automation is Changing our Living Planet](#). Routledge. This book offers a critical primer on how Artificial Intelligence and digitalization are shaping our planet and the risks posed to society and environmental sustainability.

Richard Dawkins, 1976. [The selfish gene](#). Oxford University Press. A seminal work on the basic units of biological natural selection: not the species, not the organism, but the gene. Introducing the concept of memes as basic units for cultural natural selection. Giving biological explanations of altruism.

Richard Dawkins, 1982. [The extended phenotype](#). Oxford University Press. On genes expressing themselves as the organisms that carry them (the phenotype) controlling their environment (the extended phenotype). Analogous to Hofstadter's top-down causation (memes replacing genes).

Maturana and Varela. 1972 [Autopoiesis and Cognition](#). Editorial Universitaria S.A. (1972). A generalised model of living (i.e., self-referential) systems. It was mainly applied in organisational sciences (social systems as living systems). For example, it explains why organisations synchronize their behaviour for common objectives only if a higher-order system connects them (connective leadership).

Beck & Cowan, 1996. [Spiral dynamics: Mastering Values, Leadership, and Change](#). Blackwell Publishing. Seminal work on the identification of the values that drive human behavior, called 'memes'. Distinguishing several embedded layers of consciousness, building on Graves and Maslov.

Jeroen van den Bergh 2018. [Human Evolution Beyond Biology and Culture](#). Cambridge University press. I wrote [this review](#).

Yuval Noah Harari, 2024. ['Never summon a power you can't control': Yuval Noah Harari on how AI could threaten democracy and divide the world](#). Article in the Guardian.

Podcasts (a selection)

- The rest is politics Leading podcast. 2023. [Mustafa Suleyman: Will AI save or destroy humanity?](#)
- The diary of a CEO. CEO Of Microsoft AI: AI Is Becoming More Dangerous And Threatening! - [Mustafa Suleiman](#)
- Daniel Dennett. Can we trust AI? [podcast](#). We speak with Prof. Daniel Dennett about concerns about AI being used to create artificial colleagues, he argues that preventing counterfeit AI individuals is crucial for societal trust and security. This is about his recent article in the Atlantic - "Counterfeit People."
- TED Talks daily. 2024. [The UN is speaking up about AI — here's what they're saying](#). Each Sunday, TED shares an episode of another podcast we think you'll love, handpicked for you... by us. AI is shaping every aspect of our lives — but only a handful of tech giants have a say in what this technology can do. So what's going on with world governments? Bilawal Sidhu, host of The TED AI Show, sits down with geopolitical expert Ian Bremmer to unpack the UN's just-released plan for "Governing AI for Humanity," a report that focuses on the urgent need to guide AI towards helping everyone thrive, rather than just the powerful few. Together, they explore the complexities of AI's rapid growth on a worldwide scale and take a clear-eyed look at the pivotal decisions facing us in the very near future

- TED Talks daily. 2024. [Can Europe win the age of AI? | Thomas Dohmke](#). GitHub CEO Thomas Dohmke discusses Europe's readiness to lead the next era of AI innovation, examining how the continent's tech ecosystems stack up against those in the US. In conversation with TEDAI Vienna co-curator Vlad Gozman, Dohmke explains the three key shifts that will help Europe thrive in the age of AI — and shows how GitHub's initiatives can empower anyone to build new ideas around the world.
- Closer to truth podcasts. 2024. [Nick Bostrom on Superintelligence and the Future of AI](#). Philosopher Nick Bostrom discusses his new book, *Deep Utopia: Life and Meaning in a Solved World*, where he asks: In the face of incredible technological advances, what is the point of human existence? Will AI make our life and labor obsolete? In a "solved world," where would we find meaning and purpose? Bostrom's book, *Deep Utopia*, is available for purchase now. Nick Bostrom is a Professor at Oxford University, where he is the founding director of the Future of Humanity Institute. He is the world's most cited philosopher aged 50 or under.
- Economist podcasts. 2024. [Babbage: Sam Altman and Satya Nadella on their vision for AI](#). OpenAI and Microsoft are leaders in generative artificial intelligence (AI). OpenAI has built GPT-4, one of the world's most sophisticated large language models (LLMs) and Microsoft is injecting those algorithms into its products, from Word to Windows. At the World Economic Forum in Davos last week, Zanny Minton Beddoes, The Economist's editor-in-chief, interviewed Sam Altman and Satya Nadella, who run OpenAI and Microsoft respectively. They explained their vision for humanity's future with AI and addressed some thorny questions looming over the field, such as how AI that is better than humans at doing tasks might affect productivity and how to ensure that the technology doesn't pose existential risks to society.
- Economist podcasts. 2023. [Babbage: Fei-Fei Li on how to really think about the future of AI](#). A year ago, the public launch of ChatGPT took the world by storm and it was followed by many more generative artificial intelligence tools, all with remarkable, human-like abilities. Fears over the existential risks posed by AI have dominated the global conversation around the technology ever since. Fei-Fei Li, a pioneer that helped lay the groundwork that underpins modern generative AI models, takes a more nuanced approach. She's pushing for a human-centred way of dealing with AI—treating it as a tool to help enhance—and not replace—humanity, while focussing on the pressing challenges of disinformation, bias and job disruption. Fei-Fei Li is the founding co-director of Stanford University's Institute for Human-Centred Artificial Intelligence. Fei-Fei and her research group created ImageNet, a huge database of images that enabled computer scientists to build algorithms that were able to see and recognise objects in the real world. That endeavour also introduced the world to deep learning, a type of machine learning that is fundamental part of how large-language and image-creation models work.
- Democracy IRL podcast. 2023. [How Generative AI Will Revolutionize Everything](#). Jerry Kaplan is a renowned Silicon Valley veteran, computer scientist, and serial entrepreneur who has previously authored two books on AI, with a new one on generative AI forthcoming from Oxford University Press. In this episode, he joins Francis Fukuyama to discuss why he has suddenly decided that GAI is a genuinely big deal and a technology that will fundamentally change the ways we work and live.
- TedTalks. 2019. [Kai Fu Lee on the future of AI](#). Technologist Kai-Fu Lee describes the high-stakes battle between the West and China in artificial intelligence, and what it means for the future of work.
- TEDxBerlin 2024. [Lawrence Lessig on How AI could hack democracy](#). Does AI pose a threat to democracy? Law professor Lawrence Lessig dissects how this emerging technology could influence democratic institutions, warning that we've already passed a point (before superintelligence or AGI) that deserves a lot more attention.
- Uncommon Decency podcast. 104. [Regulating AI, with Ian Bremmer & Anu Bradford](#)