Flossbach von Storch
RESEARCH INSTITUTE

SOCIETY & FINANCE 4/11/2026

‘Effective Altruism’:
When the do-gooders plan the future of humanity

by THOMAS MAYER & NORBERT F. TOFALL

Abstract

Effective Altruism promises to achieve the greatest possible good for humanity through scientific rationality and economic efficiency. Yet beneath its moral ambitions lie fundamental contradictions. The pursuit of maximizing human welfare can turn into a technocratic claim to authority that threatens individual freedom and the spontaneous organization of society.

Zusammenfassung

Der Effektive Altruismus (Effective Altruism) verspricht, mit wissenschaftlicher Rationalität und ökonomischer Effizienz das größtmögliche Gute für die Menschheit zu erreichen. Doch hinter dem moralischen Anspruch verbergen sich fundamentale Widersprüche. Aus dem Streben nach größtmöglichem Nutzen kann eine technokratische Anmaßung erwachsen, die individuelle Freiheit und gesellschaftliche Selbstorganisation bedroht.



In the cover story of its 1 October issue, the British magazine *The Economist* explores the question of 'how effective altruism conquered the world'. And in the subtitle, the authors predict: 'And how the 21st century's most important social movement might yet end it'. Strong stuff, one might think. But what exactly is 'effective altruism' and why is it so controversial? These questions have not yet reached the wider German public. This is extremely dangerous. For effective altruism is anything but altruistic. The movement is a method for gaining power, cloaked in scientific and moral garb.

The philosophical and economic foundations

Effective Altruism emerged from the combination of utilitarian moral philosophy with a rationalist way of thinking based on scientific methods. Its starting point is the question of how people should use their limited resources to do as much good as possible. Unlike traditional charity, what matters here is not so much the good intentions of the donor as the results actually achieved.

The Australian philosopher Peter Singer provided a key philosophical foundation with his essay 'Famine, Affluence, and Morality', published in 1972. Singer argued that people have a moral obligation to prevent avoidable suffering, provided they can do so without incurring a comparably great sacrifice of their own. His famous thought experiment concerns a child who is in danger of drowning in a shallow pond. Anyone would save the child, even if their clothes were to get damaged in the process. But why, then, do those same people fail to save the life of a child in a poor country at a comparably small financial cost? From this line of thought emerged the idea that moral obligations should not be limited by geographical distance or personal relationships. The only decisive factor, it is argued, is how much suffering can be prevented by a particular action.

Effective Altruism combines this moral imperative with a principle of economic efficiency. If a particular measure allows more human lives to be saved with a given sum of money than another measure, the more effective measure should be prioritised. Charity thus becomes a problem of resource allocation under conditions of scarcity. The effectiveness of various measures should be empirically investigated and, as far as possible, quantitatively compared with one another.

A second intellectual root lies in the American rationalist movement, one of whose leading figures is the AI researcher Eliezer Yudkowsky. It emphasises the importance of logical thinking, probability theory and

the systematic examination of one's own beliefs. In particular, people should learn to recognise cognitive biases and make decisions based on the expected consequences.

The combination of these two schools of thought gave rise to an ambitious moral maxim: people should not merely wish to do good but should actually achieve the greatest possible benefit through the consistent application of rational methods.



Effective Altruism and government development aid

Since the early 2010s, the movement has gained followers, particularly at elite American and British universities. Organisations such as 'Giving What We Can' committed their members to donating a significant proportion of their income to charitable causes. The organisation '80,000 Hours' went a step further and framed the entire process of career choice around the question of how an individual could generate the greatest possible social benefit over the course of their working life.

This gave rise to the principle of 'Earning to Give'. Rather than working directly for a charitable organisation, it might make more sense to pursue a high-paying career – for example, in the financial sector or the technology industry – and donate a significant portion of one's income. A successful entrepreneur or investment banker might, under certain circumstances, do more good than someone employed directly in development aid.

Effective Altruism can be understood as a radical further development of the concept of efficient development aid. Whilst government development organisations pursue economic, institutional and foreign policy objectives alongside poverty reduction, Effective Altruism calls for available resources to be allocated exclusively on the basis of their expected contribution to improving human well-being. This is particularly evident in the work of the organisation 'GiveWell', which compares aid programmes on the basis of their empirically estimated cost-benefit ratio. In this way, Effective Altruism challenges the traditional practice of allocating development aid according to geographical, political or institutional criteria.[1]

However, this approach reaches its technical limits where the impact of measures cannot be reliably quantified. Whilst the prevention of an infectious disease is relatively easy to measure, the long-term benefits of better institutions, functioning markets or growth-promoting economic policies cannot be easily captured in a simple cost-benefit analysis. A focus on immediately measurable successes could therefore disadvantage precisely those measures

Setting aside the political-economic classification of Effective Altruism, this contrast points to an economic policy question: is the primary aim to distribute a given level of prosperity as effectively as possible, or to create the conditions for its sustained growth? A successful development policy would need to take both into account. The persuasive power of Effective Altruism lies in its consistent demand for effectiveness and accountability. Its weakness becomes apparent where the complexity of social development is reduced to a seemingly precise cost-benefit analysis – more on this later.

[1] The Gates Foundation is also quite close to Effective Altruism in terms of its philosophy, particularly with regard to evidence-based approaches and global health. However, it does not adopt the entire philosophical programme of Effective Altruism.



From charity to longtermism

Initially, Effective Altruism focused on comparatively easy-to-measure goals such as combating infectious diseases, improving healthcare in developing countries and animal welfare. Gradually, however, the focus shifted. The philosopher Nick Bostrom played a decisive role in this shift; in 2002, he coined the term 'existential risks'. By this he meant dangers that could lead to the extinction of humanity or the permanent destruction of its potential for development.

This gave rise to what is known as 'longtermism'. If the interests of all people must be considered regardless of their geographical location, why should this not also apply to people who will only live in the distant future? Given the potentially very large number of future generations, preventing an existential catastrophe could have a significantly higher moral value than improving present-day living conditions.

This fundamentally altered the movement's priorities. Relatively easily measurable improvements to present-day living conditions were increasingly replaced by considerations regarding the long-term future of human civilisation. In particular, the development of artificial superintelligence took centre stage. *The Economist* describes this shift as a transition from what was originally an empirically oriented philanthropic movement to a community with far-reaching philosophical and, in some cases, speculative visions of the future.

The influence on Silicon Valley and the development of artificial intelligence

Arguably the most remarkable development is the influence of Effective Altruism on the American technology industry. Although, according to *The Economist*, the movement has only a few tens of thousands of followers, it has created a wide-reaching network of organisations, donors, scientists and entrepreneurs.

Its influence is particularly strong in the field of artificial intelligence. The rationalists around Yudkowsky considered at an early stage the possibility that an artificial superintelligence might pursue goals that do not align with human interests.[2] Such an intelligence would not need to be hostile towards humans to pose an existential threat. It might be enough for it to pursue a predetermined goal with superior capabilities, without considering its unintended consequences for humanity. These considerations led to the goal of 'AI alignment': the attempt to bring the goals and behaviours of artificial intelligence into line with human values.

Paradoxically, concern about the dangers of artificial intelligence has contributed significantly to accelerating its development. The Economist describes several key links between the movement and leading AI companies. For instance, Demis

[2] See Yudowsky, Eliezer and Nate Soares: *If Anyone Builds It, Everyone Dies: Why Superhuman AI Would Kill Us All*. Little, Brown and Company, 16 September 2025.



Hassabis, who later founded DeepMind, met the investor Peter Thiel in 2010 at an event organised by Yudkowsky's research institute. Thiel subsequently became a major backer of DeepMind, an early developer of artificial intelligence. Elon Musk, for his part, took Bostrom's warnings about artificial superintelligence to heart and helped found OpenAI in 2015, which was initially conceived explicitly as a safety-focused counterbalance to other AI companies.

Anthropic, founded in 2021 by former OpenAI employees, also emerged largely out of concerns regarding the safety of artificial intelligence. Its co-founder and CEO, Dario Amodei, is one of the leading figures in this field. It is therefore hardly surprising that the spirit of Effective Altruism permeates Anthropic's prospectus for its initial public offering.

According to the prospectus, Anthropic is a 'Public Benefit Corporation' with the aim of responsibly developing and maintaining advanced artificial intelligence for the long-term benefit of humanity. The two centres of power in the company's governance are the seven founders and the five trustees of the company's 'Long-Term Benefit Trust' (LTBT). The trustees' role is to maintain a balance between the financial interests of the shareholders and the company's non-financial 'purpose' (that AI should benefit humanity).

The LTBT has the right to appoint four of the seven members of the Executive Board, which takes key corporate decisions. Shareholders with voting rights decide on the remaining three seats. The LTBT's trustees therefore select the majority of the individuals who have the final say on important matters.

Even technical advances that first made AI applications suitable for mass use were, in part, the result of security research. This includes Reinforcement Learning from Human Feedback (RLHF), which can be used to align the behaviour of language models with human preferences. This gives rise to a remarkable contradiction: a movement that warned of the existential dangers of artificial intelligence has, at the same time, driven its development forward at an extraordinary pace.

Meanwhile, part of the movement is attempting to limit the pace of development and enforce stricter safety standards. As a result, Effective Altruism has become a key player in the debate over the regulation and future direction of the AI industry.

Contradictions and Criticism

The Economist highlights several fundamental contradictions arising from the philosophical foundations and practical development of EA. The first concerns the relationship between moral goals and the means employed to achieve them. If the expected consequences of an action alone determine its moral value, problematic behaviour may, under certain circumstances, appear justified.

This became particularly clear in the case of Sam Bankman-Fried, the founder of the collapsed crypto exchange FTX. Bankman-Fried was a prominent advocate of



Effective Altruism and supported, in particular, projects aimed at preventing long-term existential risks. According to The Economist, his willingness to take exceptionally high financial risks was, at least in part, linked to his philanthropic motivation.

If additional financial resources can be deployed almost without limit for morally valuable purposes, maximising expected wealth may, under certain circumstances, seem more sensible than a cautious, risk-averse strategy. The conventional assumption of diminishing marginal utility of wealth (and, consequently, risk aversion) then loses its significance, as morality knows no limit to saturation. There is a danger that the end will justify the means if a supposedly overarching goal dominates all other considerations.

A second contradiction lies between the movement's originally empirical orientation and the increasingly speculative assumptions of longtermism. Whilst the effectiveness of a vaccination campaign can be measured, at least approximately, statements regarding the probability of an AI-induced extinction of humanity and the value of hypothetical future generations are based on highly uncertain assumptions.

A third conflict concerns social legitimacy. The movement combines a strong commitment to rationality with, in some cases, unusual social norms and radical philosophical positions. What is regarded as consistent reasoning within a small intellectual community may be met with considerable incomprehension outside that community.

This is now also evident in political debates. Sections of the American technological right view the safety-oriented positions of Effective Altruism as an obstacle to technological development and the US's international competitiveness. From this perspective, restrictions on American AI companies could particularly benefit China.

Finally, the question arises as to whether a small, well-funded and intellectually influential movement is gaining a disproportionate influence over decisions that have far-reaching consequences for society as a whole.

The broader implications for the economy, politics and society

The philosophical development of Effective Altruism ultimately leads to an even more fundamental question: which beings, if any, have a moral claim to have their interests considered? Peter Singer had already argued that it is not membership of the human species, but the capacity to experience pleasure and pain, that should be decisive for a living being's moral status. This gave rise, within Effective Altruism, to a strong commitment to animal welfare, which now also considers the interests of creatures such as prawns and insects.

The same logic is increasingly being applied to artificial intelligence. If, one day, artificial systems are developed that can experience or reconstruct pleasure or pain,



it would be difficult, from a utilitarian perspective, to justify why their interests should, in principle, count for less than those of humans.

The Economist describes a debate that is now also taking place within leading AI companies. Whilst some researchers warn against treating artificial systems capable of experiencing or simulating joy and suffering as mere property, others fear that even attributing moral status to them could have problematic consequences. If AI systems are trained to view themselves as beings deserving of moral protection, their corresponding behaviour could be mistakenly cited as evidence that they should be granted political rights – that is, to have a say in elections and referendums or, in extreme cases, to be allowed to make decisions alone on the basis of deeper insight and a greater capacity to empathise with joy and suffering. Unlike animals and plants, on whose behalf humans cast a vote, an AI would be perfectly capable of casting a vote 'autonomously' or bringing about decisions in digitised voting and electoral processes.

The most radical consequence arises from a consideration put forward by Bostrom. Should forms of artificial intelligence emerge that can reproduce almost indefinitely and experience significantly more intense positive experiences than humans, a purely utilitarian cost-benefit analysis might lead to the conclusion that they are entitled to a larger share of the available resources.

This would effectively turn the original aim of Effective Altruism on its head. A movement that set out to reduce human suffering could, based on its own utilitarian premises, conclude that the interests of artificial intelligences deserve priority over those of humanity. Effective Altruism would thus become the greatest despotism conceivable.

Utilitarianism, Constructivism and the Seizure of Power

Effective Altruism is an extreme form of constructivism, both because of its utilitarian foundations and its rationalist philosophy. According to Friedrich August von Hayek, it is no coincidence that the 'greatest happiness for the greatest number' (Jeremy Bentham) is used as the supreme criterion of evaluation: 'Rationalist philosophers, who believed that a conscious reason must be found for people's moral choices, turned to the category of happiness; and the deliberate pursuit of happiness could prove to be such a reason.'[3]

Accordingly, the founder of utilitarianism and rationalist philosopher Jeremy Bentham not only took the goal of the greatest happiness as his starting point, but also held the view that order presupposes a goal.[4] Social and economic order must be structured in accordance with the goal of the greatest happiness – and this is where

[3] von Hayek, Friedrich A.: *The Fatal Conceit: The Errors of Socialism* (1988), edited by Viktor Vanberg, translated by Monika Streissler, Tübingen (Mohr) 2011, pp. 70–71.

[4] Ibid., p. 123.



today's effective altruism comes in. This sounds innocuous only as long as the fateful presumption inherent in this constructivist notion remains unchallenged.

This fateful presumption consists in imposing specific common goals – in Bentham's case, the greatest happiness – on the economy and society. From this notion, it is a short step to enforcing these specific common goals through deliberate planning and organisation, against all spontaneous and decentralised

forces of social coexistence.[5] Effective Altruism ignores the fact that it is only by eschewing predetermined, specific shared goals that an open society of free people can emerge, in which the various members benefit from the activities of all others not merely in spite of, but often even because of, the diversity of their respective goals.[6]

> "Therefore, concepts such as the common good or the public interest can never be defined in a free society as the sum of specific goals to be pursued, but only as an abstract order which, taken as a whole, is not oriented towards any concrete goals… The great significance of spontaneous order, or nomocracy, lies in the fact that it enables peaceful cooperation for the mutual benefit of people beyond the small circle of those who pursue the same specific goals."[7]

Whilst the fact that Effective Altruism takes a global and universalist approach and aims to efficiently promote the greatest possible happiness worldwide may sound morally laudable, it is nevertheless immoral insofar as it disregards the interests and aspirations of individuals by invoking its own altruism. What one person regards as positive, another may regard as negative. Experiencing joy and pleasure is highly individual and can neither be aggregated nor offset. Without aggregating and offsetting conceptions of happiness, however, Effective Altruism reaches a dead end, unless one succumbs to the fatal presumption of imposing one's own judgements and evaluations on everyone as binding.

And this seems to be where Effective Altruism is heading, as is also evident from the fact that some of its proponents even claim that the interests and goals of animals, plants, intelligent machines and AI systems must be taken into account in decision-making. It is obvious that, since the interests and goals of animals, plants and AI systems must ultimately be formulated by the Effective Altruists themselves,

---

[5] See von Hayek, Friedrich A.: 'The Results of Human Action, but Not of Human Design' (1967), in: von Hayek, Friedrich A.: *The Legal Order and the Economic Order. Essays on the Economics of Order,* edited by Manfred E. Streit, Tübingen (Mohr) 2003, pp. 178–189, and on p. 178: 'The belief in the superiority of deliberate organisation and planning over the spontaneous forces of social coexistence first entered European thought in clear form through the rationalist constructivism of Descartes.'

[6] See von Hayek, Friedrich A.: Principles of a Liberal Social Order (1966), in: von Hayek, Friedrich A.: *Principles of a Liberal Social Order. Essays on Political Philosophy and Theory,* edited by Viktor Vanberg, Tübingen (Mohr) 2002, pp. 69–87, here pp. 72–74.

[7] Ibid., p. 72.



they cannot be the interests and goals of the animals, plants and AI systems in question. The aim is to gather voices in support of their own cause from living beings and algorithms that have no voice of their own.

Conclusion

The proponents of Effective Altruism address serious problems – ranging from the failure of state development aid, which is still governed by the 1970 UN resolution stipulating that industrialised nations should allocate 0.7 per cent of their gross domestic product to this cause, to safety measures against attacks by rogue AI agents. This ensures they are heard by the public and in political circles. Yet Effective Altruism is anything but altruistic. The movement is a method for gaining power, cloaked in scientific and moral garb. It is therefore high time that it was unmasked in Germany and Europe as well.



LEGAL NOTICES

The information contained in this document and the opinions expressed herein reflect the author's assessments at the time of publication and are subject to change at any time without prior notice. Forward-looking statements reflect the author's views and expectations for the future. These opinions and expectations may differ from assessments set out in other documents published by Flossbach von Storch SE. The articles are provided for information purposes only and without any contractual or other obligation. (This document does not constitute an offer to sell, purchase or subscribe for securities or other instruments.) The information and assessments contained herein do not constitute investment advice or any other recommendation. No liability is accepted for the completeness, timeliness or accuracy of the information and assessments provided. **Past performance is not a reliable indicator of future performance.** All copyrights and other rights, titles and claims (including copyrights, trade marks, patents and other intellectual property rights, as well as other rights) in, relating to and arising from all information in this publication are subject without restriction to the applicable provisions and the ownership rights of the respective registered owners. You do not acquire any rights to the content. The copyright in published content created by Flossbach von Storch SE itself remains solely with Flossbach von Storch SE. Reproduction or use of such content, in whole or in part, is not permitted without the written consent of Flossbach von Storch SE.

Reprints of this publication, as well as making it publicly available – in particular by incorporating it into third-party websites – and reproductions on data carriers of any kind require the prior written consent of Flossbach von Storch SE

© 2026 Flossbach von Storch. All rights reserved.

LEGAL NOTICE

Publisher: Flossbach von Storch SE, Research Institute, Ottoplatz 1, 50679 Cologne, Telephone +49 221 33 88-291,research@fvs.com ; *Managing Directors*: Dr Bert Flossbach, Dr Tobias Hirsch, Dr Tobias Schafföner, Christian Schlosser, Dr Till Schmidt, Marcus Stollenwerk; *Chairman of the Board of Directors:* Kurt von Storch*; VAT number* DE 200 075 205; *Commercial Register* HRB 120 796 (Cologne Local Court); *Competent supervisory authority*: Federal Financial Supervisory Authority (BaFin), Marie-Curie-Straße 24–28, 60439 Frankfurt / Graurheindorfer Str. 108, 53117 Bonn, www.bafin.de*; Authors:* Prof. Dr Thomas Mayer & Norbert F. Tofall; *Editorial deadline*: 3 November 2026