ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/349821936

The ethics and governance of artificial intelligence

Article · May 2020

CITATIONS
4

READS
719

1 author:



Thorsten Jelinek
Hertie School
37 PUBLICATIONS **355** CITATIONS

SEE PROFILE

All content following this page was uploaded by Thorsten Jelinek on 05 March 2021.

The user has requested enhancement of the downloaded file.

THE ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE

Technology has been used for good or for harm, and it has fundamentally changed human relations by either extending or constraining both power and opportunity.[1] Today, the discourse on widespread digitalization and the rise of artificial intelligence (AI) amplifies both of these ethical dimensions.[2] On the upside, AI is celebrated as a new source of innovation, economic growth, and competitiveness, as well as for the productivity and efficiency improvements AI offers across all domains and sectors. Intelligent automation also promises to resolve some of the most urgent global challenges, including the urgent search for a covid-19 vaccine[3] and, more broadly, to achieve the Sustainable Development Goals.[4] The potential economic, social, and health benefits of AI innovations can be tremendous. For the majority, the rise of AI is already experienced as an increase in everyday convenience. On the downside, the rapid advancement and omnipotent nature of AI capabilities has been cautioned against and condemned as a source of unprecedented security and privacy risks as well as a source of severe social, economic, political, and international imbalances in the long term. In the past, the benefits of such dual-use or disruptive technologies eventually outweighed their harm, but this often took place only after a period of misuse and accidents that caused people and governments to demand further improvements and regulations. As AI won't be an outcome of only human agency but will increasingly develop into an independent agent of autonomous decision-making itself, we cannot readily rely on those past experiences. However, technology is not external to human behaviour, but essentially an "event to which we belong".[5] Over the next decades the main risk is not that AI itself will cause immediate harm and longterm imbalances, but our existing human intentions, relations, and actions, and thus how AI will be applied is the primary cause and source of disruption. AI won't be external to human behaviour but perpetuate and probably accelerate the current trajectory of humankind, and as history has entered a downward spiral and become more fragmented and unsustainable, the risk of experiencing more of the downside of AI is rather likely. Thus, what is urgently needed is a global mechanism for the global coordination of AI governance in order to effectively mitigate against AI-enabled cyberphysical threats and long-term structural imbalances as well as help to counter the division and fragmentation of international relations and cooperation.[6]

AI is developing fast and slow

With AI on the rise, coupled with other disruptive technologies such as 5G, the Internet of Things (IoT), robotics, quantum computing, and biosynthetics, our imaginary distance between science fiction and real science has shrunk considerably. AI already beats humans in difficult tasks like playing Chess, playing Go, playing complex strategy games, or conducting medical and legal diagnoses. Besides the intelligent

[1] A. Feenberg, *Critical Theory of Technology*, New York, Oxford University Press, 1991.
[2] For example, Stanford University, "Artificial Intelligence and Life in 2030: Report of the 2015-2016 Study Panel, Stanford University", Stanford University, Retrieved 20 June 2018, https://ai100.stanford.edu.; M. Brundage et al., *The Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation*, arXiv:1802.07228, 20 February 2018.
[3] Council of Europe, "AI and control of Covid-19 coronavirus: Overview carried out by the Ad hoc Committee on Artificial Intelligence (CAHAI) secretariat", retrieved 30 April 2020, https://www.coe.int/en/web/artificial-intelligence/ai-and-control-of-covid-19-coronavirus.
[4] R. Vinuesa, H. Azizpour et.al., "The Role of Artificial Intelligence in Achieving the Sustainable Development Goals", Nature Communications, vol. 11, no. 233, 2020, pp. 1-10.
[5] M. Blitz, "Understanding Heidegger on Technology", The New Atlantis, vol. 41, 2014, p.71.
[6] T. Jelinek, W. Wallach, D. Kerimi, "Coordinating Committee for the Governance of Artificial Intelligence", G20 Insights, 2020.

automation of control systems, computer vision and language processing have received the most attention in recent years and vastly outperform certain forms of human perception and expression.[7] Yet narrow AI is still far away from resembling general human intelligence or even reaching superhuman intelligence, and it still needs to overcome engineering bottlenecks related to creative, emotional, and social intelligence. [8] Although certain algorithms can already generate a computational sense of self,[9] today's AI is not able to efficiently abstract from one situation and apply general concepts to new contexts and tasks. Nor can algorithms automatically change the methodology of learning itself.[10] While the application of AI systems can be extremely efficient and scalable, training AI systems still takes a long time, is extremely costly, and is much more inefficient than how humans learn. From the perspective of collective intelligence, AI cannot build or compete against large and complex social organizations, which is the human ability that arguably distinguishes humankind from nature. While human consciousness provides us humans the experience of an allegedly stable reality, it's still largely unknown whether the human brain operates and represents reality based on probabilistic algorithms. In short, since the rise and collection of mass data, AI or machine learning has advanced rapidly, but it will not advance rapidly enough to match the apologetic or dystopian fantasies of a transhumanist and posthuman era anytime soon.[11] In the meantime, however, we will face tremendous challenges. As Alan Turing[12], the founding father of AI, once put it: "We can only see a short distance ahead, but we can see plenty there that needs to be done".

The changing landscape of cyber-physical threats

The level of risk attributed to AI is not a matter of optimism or pessimism but one of understanding how AI serves existing human behaviour and how it can alter power relations. Even before AI reaches or exceeds human-level intelligence, the disruptions of AI will be twofold; they will be immediate and felt directly, and they will be structural and unfold over a longer period of time.[13] Regarding the former, AI's immediate risks relate to the existing landscape of cybersecurity threats, which will change tremendously due to the malicious use of AI. There has been a steep increase in traditional cybersecurity breaches and cybercrime incidences that mainly threaten

---

[7] World Intellectual Property Organization, *WIPO Technology Trends 2019: Artificial Intelligence,* Geneva, World Intellectual Property Organization, 2019.; F. Corea, "AI knowledge map: how to classify AI technologies", Medium, 29 August 2018, retrieved 30 June 2019: https://link.medium.com/0f4BD3u754. M. Chui, J. Manyika & M. Miremadi, "What AI can and can't do (yet) for your business", McKinsey Quarterly, McKinsey & Company, January 2018.

[8] L. Floridi, "What the Near Future of Artificial Intelligence Could Be", Philosophy and Technology, vol. 32, 2019, pp. 1-15.

[9] cf. "Research Center for Brain-inspired Intelligence", Institute of Automation, Chinese Academy of Sciences, retrieved 10 November 2019, http://bii.ia.ac.cn/braincog/

[10] J. Clune, "AI-GAs: AI-generating Algorithms, an Alternate Paradigm for Producing General Artificial Intelligence", arXiv:1905.10985v2, February 2020. J. Clune, "AI-GAs: AI-generating Algorithms, an Alternate Paradigm for Producing General Artificial Intelligence", arXiv:1905.10985v2, February 2020.; S. Andersson, "Unsolved Problems in AI", AI Roadmap Institute, 3 February 2017, retrieved 10 January 2020, https://medium.com/ai-roadmap-institute/unsolved-problems-in-ai-38f4ce18921d.

[11] For example, N. Bostrom, "Transhumanist", in F. Adams (ed.), *Ethical Issues for the 21st Century,* Virginia, Philosophical Documentation Center Press, 2003.; C. Moreira & D. Fergusson, *The Transhuman Code: How to Program Your Future*, Austin, Greenleaf Book Group Press, 2019.

[12] A. Turing, "Computing Machinery and Intelligence", Oxford University Press on behalf of the Mind Association, vol. 59, no. 236, 1950, p.460. A. Turing, "Computing Machinery and Intelligence", Oxford University Press on behalf of the Mind Association, vol. 59, no. 236, 1950, p.460.

[13] Brundage, *op.cit.*

individuals, businesses, and national infrastructures. The World Economic Forum ranks cybersecurity risks among the top 5 sources of severe global risks. In the first half of 2018 cyber-attacks compromised 4.5 billion records, which was almost twice the amount of records compromised during the entire year of 2017.[14] These incidences are caused by individual criminals, organized crime groups, terrorists, and states or state-sponsored actors, and they primarily involve the disruption of digital and physical systems, theft, and cyber espionage. Cyberwarfare is a combination of all of these and also involves information and psychological operations to manipulate public opinion. Due to its scalability and efficiency as well as the increasing psychological distance between the attacker and the target, the malicious use of AI will lead to the expansion of existing cybersecurity threats, create entirely new forms of cyber-physical threats, and carry out attacks and crimes that are much more targeted by optimizing the trade-off between effort and harm or gain. Due to such changing landscape of immediate threats and risks, cybersecurity (and more recently AI) has become a matter of national security and military priority.[15] While the next generation of mobile networks, or 5G, enables to connect everything with everything and with everyone, at home, in the office, in factories, and in smart cities, AI provides automation for the purpose of efficiency and convenience. The combination of both technologies will tremendously expand the surface for cyber-physical threats and accidents. It will further complicate both the deterrence and attribution of cyber-attacks or other hacking exploits due to the increasing complexity and interconnectedness of computer networks. It won't be possible to prevent those threats, but it will only be possible to mitigate them. For many governments, it's not a question if but when severe cybersecurity incidences will occur.

Economic imbalances

In addition to these immediate risks, there are longer-term structural risks associated with AI, which are more difficult to anticipate, but their impact will be even more widespread and pervasive. This is simply because technology is not external to us, developing independently of history. Instead, it is deeply interwoven with human behaviour, and the current trajectory of humankind shows little sign of escaping from today's downward spiral of economic, societal, political, and international relations. Economically, mass labour displacement, underemployment, and de-skilling can be likely outcomes of computerisation, AI, and robotics if governments don't intervene.[16] AI directly competes with human intelligence, which was spared from automation during previous industrial revolutions. AI will not just target knowledge work but continue automating the physical labour that escaped previous waves of rationalization. As a consequence, governments must prepare for profound structural changes. Widespread automation and aging societies will reduce the labour force and labour as a major source of tax revenue. In addition, market forces have already caused the concentration of data, AI technologies, and human talents. Research and development increasingly shifts from publicly-funded to privately-owned laboratories of large AI platform companies that are less willing to share their intellectual property for the social good. While the Internet initially lowered hurdles to setting up businesses, AI raises the bar again, which can lead to digital kleptocracies and AI mercantilism if

[14] World Economic Forum, *The global risks report 2019 – 14th edition,* Geneva, World Economic Forum, 2019.
[15] European Union Agency for Cybersecurity, *ENISA Threat Landscape Report 2018*, Athens, European Union Agency for Network and Information Security, 2019.
[16] C. B. Frey & M. A. Osborne, *The Future of Employment: How Susceptible are Jobs to Computerisation?* Working Paper, Oxford Martin Programme on Technology and Employment, 17 September 2013.; for a more recent study, cf. M. Webb, *The Impact of Artificial Intelligence on the Labour Market*, Stanford University, Department of Economics, 6 November 2019, SSRN: https://ssrn.com/abstract=3482150.

the zero-marginal-cost economy remains unregulated. While rich countries will be able to afford a universal basic income for those who will not be able to re-skill, low- and middle-income countries won't be able to do the same and risk becoming trapped in their stages of development. AI coupled with data — the 'new oil' on which machine learning thrives — will disrupt the global division of labour. Countries that can't catch up with advanced automation to improve their competitiveness will be left further behind. Labour, and especially cheap labour, won't provide a sufficient comparative advantage in the future, and this will render previous development models obsolete. Income inequality has already reached alarming levels, not just between rich and poor countries but also among the rich countries. The United States has the most unequal wealth distribution among all OECD countries.[17] While a small group of transhumanists will effort and enjoy the privileges of digital upgrading, the number of those who are left behind will likely increase and add to the feeding ground for social unrest, populism, and nationalism. Before societies are able to change the meaning of labour and find new sources for improving human dignity, automation will reinforce individualism, alienation, and loneliness, and it will threaten both physical and psychological well-being and social cohesion.[18]

Political imbalances

State and political actors will make more use of AI technologies.[19] While businesses employ AI to segment people even more precisely as consumers and compete for their attention, political and state actors do so to better understand citizens as persuadable voters, followers, or potential security threats. This can help make countries more secure and the political process more efficient if AI is used responsibly and balances between economic growth, social good, and national security. However, AI increases the structural risk of shifting the power balance between the state, the economy, and society by limiting the space for autonomy. Through AI-enabled mass surveillance, psychological operations, and the weaponization of information, states and political actors might seek to acquire a disproportionate amount of power or amplify populism. The two poles of this political risk scenario are totalitarianism and tyranny of the majority. In both cases, the struggle over power dominates the struggle over progress and threatens the pillars of modern states and governments — bureaucracy, rule of law, and accountability. While authoritarian states could slide into totalitarian regimes by exerting pervasive state control and repression of differences, democracies could witness the erosion of their institutions, the polarization of their entire societies, and the disintegration of their "public morality" and "manufacturing consent".[20] Unfortunately, we can already witness the world sliding towards either pole of political imbalances, yet it has become more likely for 'liberal capitalism' developing into a 'political capitalism'.[21] AI is not the cause, but it is an increasingly weaponized tool used both within and beyond national boundaries to disrupt the political process of adversarial countries. The Edward Snowden and

[17] Organisation for Economic Co-operation and Development, *United States: Tacking High Inequalities, Creating Opportunities for All*, Paris, OECD, 2014.
[18] S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power,* London, Profile Books, 2018.
[19] A. Dafoe, "AI Governance: A Research Agenda", Working Document, v1.0, Oxford, Governance of AI Program Future of Humanity Institute University of Oxford, August 2018.
[20] E. S. Herman & N. Chomsky, *Manufacturing Consent: The Political Economy of the Mass Media,* New York, Pantheon Books, 1988.
[21] B. Milanovic, "The Clash of Capitalisms: The Real Fight for the Global Economy's Future", Foreign Affairs, January/February 2020, retrieved 13 January 2020, https://www.foreignaffairs.com/articles/united-states/2019-12-10/clash-capitalisms.

Cambridge Analytica affairs are the most known and disturbing cases of widespread cyber espionage, privacy violation, the manipulation of public opinion and interfering in the democratic process within the West. Conversely, the West frequently accuses Russia, China, North Korea, Iran, and Syria of state or state-sponsored cyber intrusions and attacks and of pervasive mass surveillance.[22]

Disrupting international relations

A fierce global competition over AI supremacy is already raging risking to disrupt existing international relations. All of the leading economies have laid out or updated their AI national strategies with the goal of promoting the development of nascent AI capabilities and ecosystems and to be able to compete globally. The Russian president, Vladimir Putin, most clearly stated the strategic importance of AI in 2017 when he said, "Whoever becomes the leader in this sphere will become the ruler of the world".[23] Russia is not leading the AI race; currently, the United States leads the race, followed closely by China. The United States wants to maintain its "leadership in AI"[24] while China aspires to become the "primary centre for AI innovation" by 2030.[25] Europe also seeks to become the "world-leading region for cutting-edge AI"[26] but it is lacking behind the United States and China in its number of AI talents and businesses, filed patents, published research papers, and investments into the AI industry for research and development.[27] All governments emphasize AI as source of growth and competitiveness. At the same time, AI is classified as a dual-use technology and is therefore subject to national security, export controls, and FDI screening mechanisms. Governments have hastily passed new regulations to mitigate cybersecurity risks, ensure privacy protection, and empower law enforcement. The new regulations also protect domestic markets under the banner of digital and data sovereignty.[28] The head-to-head race has extended to national defence agencies who are preparing for a "hyperwar"[29] and making "battlefield-ready AI"[30] a priority. Most troubling of all is the development of lethal autonomous weapons (LAW). While the European Union is calling for a ban of 'automated killing robots,' the United States, China, Russia, and other countries are all advancing or acquiring LAW capabilities. Compared to conventional weapons, cyber weapons are low-cost and more easily accessible,

---

[22] Q. E. Hodgson et al., *Fighting Shadows in the Dark: Understanding and Countering Coercion in Cyberspace*, Santa Monica, Calif. RAND Corporation, 2019.
[23] "'Whoever leads in AI will rule the world': Putin to Russian children on Knowledge Day"; RT, 1 September 2017, retrieved 20 February 2020, https://www.rt.com/news/401731-ai-rule-world-putin.
[24] Executive Order 13859, *Maintaining American Leadership in Artificial Intelligence,* Executive Office of the President, Federal Register of the United States Government, Washington DC, 11 February 2019.
[25] State Council, 新一代人工智能发展规划 35 号 ["*Next generation artificial intelligence development plan No. 35* "], The State Council, People's Republic of China, 20 June 2017.
[26] Commission of the European Communities, "Member States and Commission to work together to boost artificial intelligence 'made in Europe'", Press Release, Brussels, 7 December 2018.; Commission of the European Communities, *On Artificial Intelligence - A European Approach to Excellence and Trust*, COM(2020) 65 final, Brussels,19 February 2020.
[27] Tsinghua University, *China AI Development Report 2018*, China Institute for Science and Technology Policy, Tsinghua University, Beijing, 2018.; P. Lorenz & K. Saslow, *Demystifying AI & AI companies: What foreign policy makers need to know about the global AI industry,* Stiftung Neue Verantwortung, Berlin, 2019.
[28] cf. Lorenz & K. Saslow, *op.cit.*
[29] A. Husain et al., *Hyperwar: Conflict and Competition in the AI Century,* Austin, SparkCognition Press, 2018.
[30] T. Ashizuka, "Pentagon seeks to triple AI warfare budget to meet China's rise", Nikkei Asian Review, 4 October 2019, retrieved 20 January 2020, https://asia.nikkei.com/Business/Aerospace-Defense/Pentagon-seeks-to-triple-AI-warfare-budget-to-meet-China-s-rise.

which will accelerate the diffusion of cyberwarfare and LAW capabilities. This will also empower otherwise weaker actors, thus tremendously increasing the risk of asymmetric conflicts.[31] Due to the proliferation of cyber technologies and the ongoing rush by many states to obtain offensive cyber capabilities for potential use in conflict, the actual risk of international cyberconflict and cyberwarfare has increased significantly, that is using use digital technology by one country to disrupt vital digital systems of another country. Such proliferation of technologies also holds the risk of "friendly fire" and "second order consequences" because many cyber networks rely on some private sector infrastructure. For defence agency, AI is not about good or harm, but about competition and conflict.

There are numerous international organizations dealing with cybersecurity and cybercrime, but the realm of AI-enabled cyberconflict lacks international treaties and attempts to build familiarity, mutual trust, and confidence, especially between rival powers. [32] The discourse on cybersecurity and cybercrime prevention is globally fragmented, and conventional arms-control treaties are being ripped apart or put into question.[33] In addition, the United States is trying everything it can to decouple its technology, research, and supply chains from that of China, and it is pushing Europe and its other allies and partners to do the same. The United States is doing this to confine China's rise based on national security concerns, though it has failed to provide evidence of misconduct. The clash between the United States and China has been most obvious on the issue of 5G, and it has increasingly extended to other disruptive technologies such as AI, IoT, robotics, quantum computing, and biosynthetics.[34]

The prospects for AI ethics and governance

While we cannot anticipate the outcome of the digital and AI revolution because history gives us little to no reference for what could be the final technological revolution, such sobering lists of immediate threats and longer-term structural imbalances and tensions have sparked an international debate about the ethics and governance of AI. In this debate, the term ethics is often used to summarize those legitimate concerns about these potential disruptions of AI. The debate about AI ethics and governance has resulted most notably in the definition of numerous AI principle frameworks worldwide, which have been primarily proposed by large Internet platforms and multinational corporations, as well as international and non-governmental organizations and governments.[35]

Despite subtle but crucial differences in selecting and emphasizing certain ethical principles, the various principle frameworks commonly emphasize that future AI should be secure, safe, explainable, fair, and reliable, and they also emphasize that its benefits should be distributed across society. There seems to be an international consensus that AI should be developed and used for the greater good of humanity. It

[31] A. Polyakova, "Weapons of the Weak: Russia and AI-driven Asymmetric Warfare", A Blueprint for the Future of AI, The Brookings Institution Artificial Intelligence and Emerging Technology, 2018.
[32] Brundage, *op.cit.*
[33] "U.S. Suspends Nuclear Arms Control Treaty with Russia", New York Times, 1 February 2019, retrieved 30 May 2019, https://www.nytimes.com/2019/02/01/us/politics/trump-inf-nuclear-treaty.html.
[34] J. L. Schoff & A. Ito, *Competing with China on technology and innovation*, Alliance Policy Coordination Brief. Washington, Carnegie Endowment for International Peace. October 2019.
[35] Y. Zeng, E. Lu & C Huangfu, *Linking Artificial Intelligence Principles,* arXiv:1812.04814, 12 December 2018.; A. Jobin, M. Ienca & E. Vayena, *The Global Landscape of AI Ethics Guidelines*, arXiv:1906.11668, 24 June 2019.

should be human-centric, responsible and trustworthy, and it should always retain human agency and human oversight. Yet this positive framing primarily confirms, conversely put, that today's ethics and governance are ill-equipped to prevent or sufficiently mitigate the disruptive forces of AI and that those potential forces are clearly of global and historical proportions. However, almost all frameworks analyse the risk of AI in a narrow sense: that is, without developing a link between the dual-use character of the technology and the actual state of social, political, economic, and international affairs.[36] Those frameworks ignore how AI will most likely reinforce rather than alter the current trajectory of history as indicated above. AI will increasingly make autonomous decisions, but it won't escape and be completely autonomous from human practices anytime soon and we cannot expect it to become a transcendent, super-beneficial, and human-centric compass directing humanity toward universal equality and dignity. While many of these AI principles were quickly defined, the definition of new governance approaches, that are supposed to implement those principles, will be necessary but more difficult given AI's complex and uncertain risk scenario.[37]

Governance is the possibility for collaboration directed by common principles.[38] Collaboration is necessary, as each stakeholder faces different responsibilities and no stakeholder alone can tackle AI risks in their entirety.[39] However, fundamental political and cultural differences especially between the major economic blocs undermine international collaboration.[40] Even so, collaboration and cooperation will become more urgent in the future to effectively address the risks of AI. Those fundamental differences make the looming ethics and governance gap seemingly insurmountable.[41] Accordingly, the United States is a market foundationalist economy and individualist society following the motif of profit and personal self-fulfilment. The government emphasizes AI as an opportunity for R&D, growth, and job creation. While cybersecurity risks are treated as a liability, AI mainly serves the capitalist ethics of self-directed wealth creation. In contrast, Europe emphasizes solidarity and a human rights approach to AI. According to the European Union, AI should be lawful, robust, and ethical. The mitigation of AI risks is a matter of regulation and intervention. In China, harmony and compassion are emphasized as the country's overarching virtues. For China's government, data and AI are a means of maintaining or improving stability and discipline through surveillance and control. While Chinese people largely perceive the digital revolution as an opportunity, Western people tend to emphasize the dangers of widespread digitalization. Those differences face and most likely reinforce an already divided and fragmented cyber regime complex.[42]

Undoubtedly, such polarization risks to omit the many differences within each region and the similarities across all the regions. People in Europe, the America, and China have become increasingly aware of the privacy and security risks related to

[36] Feenberg, *op.cit.*
[37] A. Dafoe, "AI Governance: A Research Agenda", Working Document, v1.0, Oxford, Governance of AI Program Future of Humanity Institute University of Oxford, August 2018.
[38] Ibid.
[39] W. Wallach & G. E. Marchant, "Toward the Agile and Comprehensive International Governance of AI and Robotics", Proceedings of the IEEE, vol. 107, no. 3, 2019, pp. 505-508.
[40] T. Jelinek, "Post Rapprochement: Sino-West Relations", in W. Billows & S. Körber (eds.), *Cultures of We? Europe and the Search for a New Narrative*, European National Institutes for Culture and Institut für Auslandsbeziehungen (ifa), Göttingen, Steidl, 2018, pp. 54-59.
[41] A. Daly et al., *Artificial Intelligence Governance and Ethics: Global Perspectives*, arXiv:1907.03848, June 2019.
[42] J. S. Nye, *The Regime Complex for Managing Global Cyber Activities*, Belfer Center for Science and International Affairs, Harvard Kennedy School, Cambridge, Massachusetts, 2014.

ubiquitous digitalization and AI. Governments have hastily sought to create a balance between security and autonomy with the goal to harness the benefits while simultaneously minimizing the risks. Large Internet and AI platforms have been pushed to become more responsible. The big powers face the same challenges, yet they approach them from different ends, which undermines the prospect of international collaboration and governance also in the area of AI. The differences are firmly rooted in the history and cultures of the regions, and they have recently become amplified. The United States and China have especially lost patience with trying to understand one another after a long period of rapprochement.[43] Instead, they forcefully articulate and defend their otherness. China perceives the world as a "community of shared destiny," while the United States emphasizes competitive coexistence as the new foundation of their relationship.[44] The former evokes a utopian harmony; the latter is more realistic, but it sparks tension and lacks a concept for peaceful coexistence.

As a result, today's global context brings us dangerously close to a never-ending pre-war scenario between China and the United States. Both powers are pushing towards the "Thucydides Trap".[45] The past globalism of the 1990s and 2000s threatens to turn into a post-global reality, one of competing national globalists repeatedly failing to reach a consensus for the development of a new equilibrium and multilateral order.[46] The disintegration of the World Trade Organization and erosion of the old United States-led order brings us back to an era where 'might is right'. It is an era of allegiances and fragmented bilateralisms. It is an era of high uncertainty and seemingly uncontrollable risks, where many have lost trust in businesses, technology, and local and global institutions, certainly within the West. Europe has become more 'real'. Yet the 'European awakening' remains a precarious one as the region will continue balancing between breaking up, heightened xenophobia, sluggish growth, and protecting the 'European way of life' but without the capacity for global stewardship. Like the United States, Europe has yet to find an escape path from the growing rift between its "brahmin left" and "merchant right".[47] Like the United States, Europe fails to represent the struggles and anxieties within its societies. Europe remains sandwiched between a "protectionist" United States, an "aggressive" China, and the rivalry between the two countries.[48] While Europe largely disagrees with President Trump's personality and approach, it shares the complaints and concerns about China's increasing dominance and China's lack of turning more Western. The European Union has also started labelling China a "strategic rival," but without joining the United States' unilateral trade war and geopolitical aggression against China. Although the United States seems to fear its future the most, China must also try harder to find a way to reduce such fear. For now, further harm is only prevented as each of the three powers is an important trading partner of the other two.

---

[43] Jelinek, *op.cit.*

[44] A. Ortega, *The U.S.-China race and the fate of transatlantic relations. Part 1: Tech, Values, and Competition*, Center for Strategic and International Studies, 13 January 2020.

[45] G. Allison, *Destined for War: Can America and China Escape Thucydides' Trap?* Wilmington, Houghton Mifflin Co, 2017. G. Allison, *Destined for War: Can America and China Escape Thucydides' Trap?* Wilmington, Houghton Mifflin Co, 2017.; A. Waldron, "There is no Thucydides Trap", SupChina,12 June 2017, retrieved 30 October 2019, https://supchina.com/2017/06/12/no-thucydides-trap.

[46] J. Morse & R. O. Keohane, "Contested multilateralism", The Review of International Organizations, vol. 9, 2014, pp.385–412.

[47] T. Piketty, *Capital and Ideology,* Cambridge, Massachusetts, Harvard University Press, 2020.

[48] J. Brunsden, S. Fleming & A. Beattie, "EU ill-equipped to face China and US, Brussels trade chief warns", Financial Times, 9 October 2019, retrieved 10 October 2019, https://www.ft.com/content/d5d1e928-e9c6-11e9-a240-3b065ef5fc55.; M. Esteban & M. Otero-Iglesias (eds), *Europe in the face of US-China rivalry,* A Report by the European Think-tank Network on China (ETNC), January 2020.

Outlook

Against such hyperbolized backdrop, it becomes obvious that AI will mainly be used to gain a strategic advantage over other competitors and rivals. Like capitalism, AI is disruptive and lacks the ethics of social good. Therefore, breaking through the current downward spiral and ensuring that technology is primarily used for social and ecological good remains a matter of human agency, collaboration, and cooperation. AI bears its own risks. However, AI is not the primary cause of changing history but a technology with a high risk of intensifying the symptoms of history, such as the unequal attention economy, the surveillance state, and greater power competition. To break through the downward spiral and to prevent an ideologization of the various AI principles, AI doesn't merely need to serve humanity; humanity and global politics needs to change as well. If human history continuously faces backlashes, then AI should not mimic human values and the human brain but learn from and improve both. If we embark upon the assumption that humanity is a threat to life on Earth and the goal is to reverse that threat, AI will have a chance to eventually improve human relations and our relationship with nature.

Therefore, the most powerful countries, which are still best represented by the Group of Twenty (G20), must come together and make the responsible governance of AI a priority. As technology itself will not resolve humanity's global challenges and history has entered a critical path, governments must agree upon a global mechanism to plan and coordinate for the responsible governance of AI.[49] Such mechanism would be instrumental in mitigating direct cyber-physical threats as well as longer-term structural imbalances. Such mechanism would also help to counter an already fragmented digital regime complex and, conversely, strengthen an emerging polycentricity and governance networks of AI that remain best suited for responding to local requirements. More broadly, a global mechanism for coordinating the governance of AI can become a benchmark for tolerating and managing ambiguity and conflict. Governments must look for common views, respect differences, and facilitate a reasonable debate over differences in hopes of forging common views over time.

BIO

Dr. Thorsten Jelinek is a senior research fellow and the European director of the Taihe Institute. His research and international advisory work focus on ethics, governance, and security-related questions concerning artificial intelligence, 5G, and digitalization. Dr. Jelinek was an associate director at the World Economic Forum and worked in management positions for leading technology firms. He is a member of the T20/G20 taskforce on the Future Multilateralism and Global Governance, a member of OECD's Trust in Business Initiative, and participates at the UN High-Level Panel on Digital Cooperation. He holds a PhD in social and political sciences from the University of Cambridge, a MSc in organizational and social psychology from the London School of Economics, and an MA in software engineering and business administration from the Beuth University of Applied Sciences in Berlin.

[49] Jelinek, Wallach, Kerimi, *op.cit.*

View publication stats