

INTERNATIONAL COOPERATION ON ARTIFICIAL INTELLIGENCE- A PROBLEM STATEMENT

The following contribution sheds light on the norms at stake when it comes to international cooperation on artificial intelligence (AI). International cooperation on AI involves two conceptual pillars: international regulation and global diffusion of AI-powered technologies. While the international regulatory discourse is concerned with establishing frameworks for minimizing and mitigating negative consequences of AI development and deployment, the global diffusion of AI technologies presents global policy makers with an ethical dilemma. AI represents an important tool to realize social development, but also an instrument that can amplify pre-existing norm violations. Based on an analysis of the moral dilemmas posed by international cooperation on AI, the authors advocate focusing on the avoidance of complicity in human rights violations while maintaining technology transfer of AI solutions critical for human development.

Alexander Kriebitz* & Christoph Lütge**



* Alexander Kriebitz: Postdoctoral Researcher at Technical University of Munich.

** Christoph Lütge: Chair of the Peter Löscher Chair of Business Ethics and Director of the Institute of Ethics in Artificial Intelligence.

International cooperation on artificial intelligence (AI) has been hailed by international organizations, multinationals developing and deploying AI, universities and regulatory bodies. Moreover, international AI cooperation encompasses two key facets, namely, the international regulation of AI and its global diffusion.¹ In response to intensifying collaboration in both areas, human rights scholars and policy makers have voiced concerns over AI collaboration with states accused of violating international norms.² Based on human rights and national interest considerations, foreign policy makers have advocated measures to reduce international cooperation on AI, particularly with countries regarded as authoritarian.³ Owing to this dissonance between emerging AI cooperation and efforts to curb the global diffusion of AI, the following contribution sheds light on the norms at stake when it comes to international cooperation on AI. The paper articulates the ethical problem statement of international cooperation on AI pertaining to the two following questions:

- What aspects do organizations need to consider when interacting commercially with actors outside of the OECD?
- Is it legitimate to engage in technology transfer on AI with authoritarian regimes?

The remainder of the publication is structured as follows: The paper begins with a review of the existing discourse on international technology cooperation to delineate potential areas of norm collisions that might be relevant for AI. In a second step, the contribution concentrates on the impact analysis of AI on the discussed norms. On this basis, the authors examine the prospective role of proportionality as a principle to reconcile conflicting norms and interests.⁴

Normative Concepts in International Affairs

In the following contribution, we understand “normative” as a general property pertaining to morally desirable aims, situations, or actions.⁵ Normative considerations

¹ World Bank. *Better Data for Doing Good: Responsible Use of Big Data and Artificial Intelligence*, OECD/LE-GAL/0449, (2018).

² A. Polyakova and C. Meserole, *Exporting Digital Authoritarianism: The Russian and Chinese Models. Policy Brief, Democracy and Disorder Series* (Washington, DC: Brookings, 2019): p. 1-22; G. King, J. Pan and M. E. Roberts, “How Censorship in China Allows Government Criticism but Silences Collective Expression,” *American Political Science Review*, Vol. 107, No. 2 (Month 2013): p. 326-343.

³ H.R. 3426 (IH) - *Democracy Technology Partnership Act*.

⁴ S. S. ÓhÉigeartaigh et al., “Overcoming Barriers to Cross-Cultural Cooperation in AI Ethics and Governance,” *Philosophy & Technology*, Vol. 33, No. 4 (Month 2020): p. 571-593.

⁵ C. Lütge, *Order Ethics or Moral Surplus: What Holds a Society Together?* (Lexington Books, 2015); C. Lütge and M. Uhl, *Business Ethics: An Economically Informed Perspective* (USA: Oxford University Press, 2021).

concern various aspects of AI cooperation beyond AI regulation, namely, the diffusion of AI solutions on a global level by investment, finance, trade, or development aid. These modes of cooperation are international in nature and hence fall under the domain of international ethics. Thus, the UN Charter commonly understood as the gravitational center of this normative space develops programmatic goals the United Nations ought to realize in Article I.⁶ Moral desiderata of international affairs include international peace (UN Charter, Art. 1[1]); the solution of “international problems of an economic, social, cultural, or humanitarian character” (UN Charter, Art. 1[3]), and “respect for human rights” (UN Charter, Art. 1[3]). International treaties have further specified the aims articulated in Article I, including the Universal Declaration of Human Rights (1948), the International Covenant of Political and Civic Rights (1967), and the International Covenant of Economic, Social and Cultural Rights (1967). International AI cooperation is situated within this comprehensive normative corpus.

“International cooperation deals with the misuse of technologies by engaging against cybercrime or preventing distribution of weapons of mass destruction to terrorist organizations.”

The Normative Debate on Technology Cooperation

Technologies matter for the (non-)realization of moral aims in international affairs, creating significant repercussions for individuals, collective organizations, or the planet as a whole.⁷ Besides, technologies produce intended and unintended consequences.⁸ Unintended consequences may be disasters caused by technical error, human miscalculations, or ignorance of long-term consequences – including “ultra-hazardous activities”; intended harm pertains to the deliberate use of technologies at the expense of the rights and interests of others.⁹

Consequently, international cooperation matters for harnessing technology in a way

⁶ A. Verdross, “General International Law and the United Nations Charter,” *International Affairs* (Royal Institute of International Affairs 1944-1954): p. 342-348; S. Chesterman, I. Johnstone and D. Malone, *Law and Practice of the United Nations: Documents and Commentary* (Oxford: Oxford University Press, 2016).

⁷ H. Jonas, *Toward a Philosophy of Technology* (Hastings Center Report, 1979): p. 34-43.

⁸ R. N. Osborn and D. H. Jackson, “Leaders, Riverboat Gamblers, or Purposeful Unintended Consequences in the Management of Complex, Dangerous Technologies,” *Academy of Management Journal*, Vol. 31, No. 4 (1988), p. 924-947.

⁹ A. E. Boyle, “Globalising Environmental Liability: The Interplay of National and International Law,” *Journal of Environmental Law*, Vol. 17, No. 1 (2005): p. 3-26.

that does not threaten but enhances major aims in international affairs. This importance expresses itself in various channels and formats of collaboration: International frameworks and procedures are intended to mitigate risks of technologies, when human and technical error might cause immense damage.¹⁰ In addition, interdependencies render international regulation necessary when it comes to basic traffic rules, the mitigation of natural catastrophes by early warning systems, or collaboration in space. Likewise, international regulation aims at incentivizing technological progress through intellectual property rights (WIPO) or frameworks enabling international competition. International cooperation deals with the misuse of technologies by engaging against cybercrime or preventing distribution of weapons of mass destruction to terrorist organizations.¹¹ Thus, international cooperation involves norm enforcement and creation. Based on the latter pillar, the international community has formulated legal provisions against biological weapons, human cloning, or lasers to blind human beings.¹²

The Necessity of International Technology Transfer

Technology transfer depicts a quintessential aspect of international technologies cooperation as it pertains to the global diffusion of technologies, skills, and knowledge.¹³ A strong argument for international technology transfer emerges from the UN Charter outlining the purpose of international cooperation in terms of “solving international problems of an economic, social, cultural, or humanitarian character” (UN Charter Art. 1[3]).

Economic imbalances between developing and developed nations depict one of the most pressing issues in contemporary international affairs, manifesting in different levels of life expectations, education, growth rates and literacy or access to medical treatment and food.¹⁴ The normative aim to realize a fair distribution of global capabilities, resources, and wealth manifests itself in the notion of a “right to development.” The Declaration on the Right to Development adopted by the UN

¹⁰ S. Miller, *Dual Use Science and Technology, Ethics and Weapons of Mass Destruction* (Dordrecht: Springer, 2018): p. 2018.

¹¹ G. Calcara, “The Role of INTERPOL and Europol in the Fight Against Cybercrime, with Reference to the Sexual Exploitation of Children Online and Child Pornography,” *Masaryk University Journal of Law and Technology*, Vol. 7, No. 1 (2013): p. 19-33.

¹² M. H. Arsanjani, “Negotiating the UN Declaration on Human Cloning,” *American Journal of International Law*, Vol. 100, No. 1 (2006): p. 164-179; J. Goldblat, “The Biological Weapons Convention: An Overview,” *International Review of the Red Cross (1961-1997)*, Vol. 37, No. 318 (1997): p. 251-265; J. H. McCall Jr, “Blinded by the Light: International Law and the Legality of Anti-Optic Laser Weapons,” *Cornell Int’l LJ*, Vol. 30, No. 1 (1997).

¹³ J. Gottwald, L. F. Buc and W. Leal Filho, “Technology Transfer,” in S. O. Idowu, N. Capaldi, L. Zu, and A. D. Gupta (eds.), *Encyclopedia of Corporate Social Responsibility* (Berlin, Heidelberg: Springer, 2013). https://doi.org/10.1007/978-3-642-28036-8_673

¹⁴ A. Sengupta, “On the Theory and Practice of the Right to Development,” *Human Rights Quarterly*, Vol. 24, No. 4 (2002): p. 837-889.

General Assembly in 1986 refers here to “equality of opportunity for all in their access to basic resources, education, health services, food, housing, employment and the fair distribution of income” (Art. VIII [1]).¹⁵ The postulated right encapsulates the idea that “nations have a duty to contribute to the assimilation of living standards internationally.”¹⁶ Thus, literature suggests that basic innovations play a critical role in realizing similar living standards in the international context and as an instrument to realize long-term growth.¹⁷ Moreover, the notion of inequalities based on access to technologies has been articulated in the term “global digital divide,” implying that access to certain technologies characterizes different living standards.¹⁸ Moreover, the International Covenant on Economic, Social and Cultural Rights defines rights (including rights to health or food) as concrete aims of national and international efforts. Denying developing countries access to technologies in critical areas such as technologies for vaccine distribution would constitute human rights violations.¹⁹

“The right to development and human rights in their social and economic dimensions correspond to the interest of the global majority in establishing global minimum standards. Here, decision makers should refrain from unilateral measures enhancing preexisting technological divides.”

The Discontents of International Technological Cooperation and Technology Transfer

In spite of the prospective impact of technology on social development, international observers have voiced concerns over technology transfer on moral grounds.²⁰ This applies specifically to the view that technologies might threaten the realization

¹⁵ I. D. Bunn, “The Right to Development: Implications for International Economic Law,” *Am. U. Int’l L. Rev.*, Vol. 15, No. 1425 (1999).

¹⁶ A. Sengupta, “Implementing the Right to Development,” *International Law and Sustainable Development* (Brill Nijhoff, 2004): p. 341-377.

¹⁷ R. M. Solow, “Technical Change and the Aggregate Production Function,” *The Review of Economics and Statistics*, (1957): p. 312-320; R. M. Solow, “A Contribution to the Theory of Economic Growth,” *The Quarterly Journal of Economics*, Vol. 70, No.1 (1956): p. 65-94.

¹⁸ W. Chen and B. Wellman, “The Global Digital Divide—Within and Between Countries,” *IT & Society*, Vol. 1, No. 7 (2004), p. 39-45.

¹⁹ U.N. CHR, General Comment 31, “The Nature of the General Legal Obligation Imposed on States Parties to the Covenant,” Article 2, (29 March 2004), CCPR/C/74/CRP.4/Rev.6.

²⁰ D. Scissors and S. Bucci, “China Cyber Threat: Huawei and American Policy Toward Chinese Companies,” *The Heritage Foundation* (2012): p. 3761; Polyakova and Messerole (2019).

of international peace, human rights, or national interest. This belief originates in historical perceptions: Nuclear weapons, cyberwars, and militaries have reconfigured international power structures, and technologies might empower actors in international affairs with an agenda adverse to human rights or international peace. This concerns not only diffusion of lethal weapons, but also the spread of dual-use items and fundamental science. Likewise, technologies can bear major ramifications for situations inside of authoritarian regimes when directed against individual rights. Consequently, literature in the business and human rights discourse embarked on exploring the implications of the potential misuse of technology transfer by authoritarian leaders.²¹ The Guiding Principles on Business and Human Rights articulates here the view that business operations “should address adverse human rights impacts with which they are involved.” Efforts to limit technology transfer originate in different ideas: preventing direct complicity in human rights/humanitarian law violations and containing geopolitical adversaries.

International Cooperation on AI as a Moral Dilemma

The discourse on technology transfer confronts decision makers with conflicting implications: Technologies require a set of global minimum standards, their diffusion is vital for the convergence of developing nations, and they might amplify preexisting tendencies toward norm violations. Consequently, the remainder of the paper is focused on the prospective role of AI within this context.

The Need for International Regulation of AI

AI combines large amounts of data with intelligent algorithms, allowing the software to learn automatically from given data input.²² A further communality of AI solutions is based on their capability to imitate intelligent human behavior.²³ Consequently, literature defines AI as a “family of technologies,” implying that AI-based human rights violations may have different causes and origins.²⁴ Specifically, AI can trigger unintended consequences—for example, biases may trigger discrimination, or organizations and agents can use AI with the explicit intention to create damage.²⁵

²¹ UN Guiding Principles on Business and Human Rights Principle 11.

²² Compare: EU AI Act; R. S. Michalski, J. G. Carbonell and T. M. Mitchell (eds.), *Machine Learning: An Artificial Intelligence Approach* (Springer Science & Business Media, 2013).

²³ A. Kaplan and M. Haenlein, “Siri, Siri, In My Hand: Who’s the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence”, *Business Horizons*, Vol. 62, No. 1 (2019): p. 15-25;

C. Bartneck, C. Lütge, A. Wagner and S. Welsh, *An Introduction to Ethics in Robotics and AI* (Springer Nature, 2021) p. 117.

²⁴ A. Kriebitz and C. Lütge, “Artificial Intelligence and Human Rights: A Business Ethical Assessment,” *Business and Human Rights Journal*, Vol. 5, No. 1 (2020): p. 84-104; Bartneck, Lütge, Wagner, and Welsh (2021), p. 117.

²⁵ A. Lambrecht and C. Tucker, “Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads,” *Management Science*, Vol. 65, No. 7 (2019): p. 2966-2981.

Consequently, international cooperation needs to cover both aspects. The black-box character of AI and questions linked to the human oversight–responsibility nexus necessitate common guidelines and codes of conduct as minimum standards for AI development and deployment. Interdependency constitutes a further variable in the equation, determining the need for international AI cooperation. This applies in specific to the international data transfer. The purpose of such cooperation would be to render AI technologies safer and better suited for realizing moral aims by establishing principles for a global AI ecosystem and outlawing misuse of AI, such as subliminal techniques, manipulative technologies, or certain supervision technologies.²⁶ A precondition of this regulatory discourse is to comprehend exactly how AI affects collective aims.

AI as an Enabler of Collective Aims

Preceding literature illustrates that artificial intelligence is conducive to the realization of human rights, social development, and economic growth.²⁷ Given its economic relevance, studies have warned that an imbalanced distribution of AI could exacerbate preexisting inequalities on a global level. Future differences in living standards could therefore express themselves in how societies adapt and use artificial intelligence. Therefore, it is plausible to connect the notion of “right to development” with policies incentivizing a global diffusion of vital AI technologies that safeguard a global minimum of AI deployment.²⁸ Moreover, AI solutions often address particular aspects relevant to human rights realization. This applies specifically to the right to health, where researchers highlight a range of technologies conducive to global health.²⁹ For example, AI solutions could help to detect tuberculosis or cervical cancer.³⁰ Furthermore, autonomous driving depicts a basic technology enhancing passenger and traffic safety.³¹ Consequently, collaboration on AI is of specific importance for resource-poor countries.³²

²⁶ World Health Organization, *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*, (2021).

²⁷ Y. He, “The Importance of Artificial Intelligence to Economic Growth,” *Korea Journal of Artificial Intelligence*, Vol. 7, No. 1 (2019): p. 17-22; J. Bughin et al., *Artificial Intelligence the Next Digital Frontier?* (McKinsey Global Institute, 2017); PwC, *The Macroeconomic Impact of Artificial Intelligence* (2018).

²⁸ A. Korinek and J. E. Stiglitz, *Artificial Intelligence, Globalization, and Strategies for Economic Development*, (No. w28453), (National Bureau of Economic Research, 2021).

²⁹ A. Ismail and N. Kumar, “AI in Global Health: The View from the Front Lines,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (May 2021): p. 1-21.

³⁰ N. Schwalbe and B. Wahl, “Artificial Intelligence and the Future of Global Health,” *The Lancet*, Vol. 395, No. 10236 (2020): p. 1579-1586.

³¹ C. Lütge, “The German Ethics Code for Automated and Connected Driving,” *Philosophy & Technology*, Vol. 30, No. 4 (2017): p. 547-558.

³² B. Wahl et al., “Artificial Intelligence (AI) and Global Health: How Can AI Contribute to Health in Resource-Poor Settings?,” *BMJ Global Health*, Vol. 3, No. 4 (2018), e000798.

AI as a Threat to Collective Aims

As mentioned above, AI represents a comprehensive set of technologies, with certain AI solutions contravening human rights and international law. AI-driven supervision technologies, such as certain manipulative AI technologies or the use of AI-powered lie detectors or social credit ratings, have repeatedly sparked ethical concerns due to their invasiveness of privacy and related to freedom of expression.³³ A further area of concern is the use of neutral technologies (such as face recognition) for realizing controversial aims. This applies to settings in which powerful actors have interests in using AI to enhance their control over others.³⁴ In particular, Western NGOs and scholars have scrutinized a wide range of countries, including Angola, Ethiopia, China, Russia, and Venezuela concerning developing, using, or exporting AI-driven technologies in ways that infringe upon human rights.³⁵ Nevertheless, similar incentive structures might prevail beyond countries classified as authoritarian, as actors in democracies could be tempted to (mis-)use similar technologies in moments of crisis.³⁶ Similar arguments have been raised in the military context of AI: AI might not only be used for upgrading conventional weapons, but also to coordinate cyber-attacks or as robot soldiers. Scholars have warned that AI-based robots would have difficulties distinguishing between civilians and combatants, enhancing risks of breaches of the Geneva Convention.³⁷ Consequently, uncontrolled AI diffusion could have adverse impacts on human rights and humanitarian international law, owing to its nature of standardizing processes and reducing human oversight.

A Sketch for Ethical AI Cooperation

Comparing the technology transfer discourse and the debate on AI ethics reveals the moral dilemma decision makers face when dealing with international cooperation and technology transfer. Thus, AI raises the stakes in the pre-existing collision between different norms and interests of international cooperation. In the following, we sketch a set of preliminary suggestions derived from the parallelism of technology transfer and AI ethics discourses.

³³ E. Donahoe and M. M. Metzger, "Artificial Intelligence and Human Rights," *Journal of Democracy*, Vol. 30, No. 2 (2019): p. 115-126.

³⁴ A. Kriebitz and R. Max, "The Xinjiang Case and Its Implications from a Business Ethics Perspective," *Human Rights Review*, Vol. 21, No. 3 (2020): p. 243-265.

³⁵ J. Millward and D. Peterson, *China's System of Oppression in Xinjiang: How it Developed and How to Curb it* (Brookings Institution, 2020); A. Polyakova and C. Meserole, *Exporting Digital Authoritarianism: The Russian and Chinese models* - Policy Brief, Democracy and Disorder Series (Washington, DC: Brookings, 2019): p. 1-22.

³⁶ P. Molnar, *Robots and Refugees: The Human Rights Impacts of Artificial Intelligence and Automated Decision-Making in Migration – Research Handbook on International Migration and Digital Technology* (Edward Elgar Publishing, 2021).

³⁷ U. Pagallo, "Robots of Just War: A Legal Perspective," *Philosophy & Technology*, Vol. 24, No. 3 (2011): p. 307-323.

Mitigating Unintended Consequences of AI

Irrespective of the particular interests of AI developers or users, all sides share the aim of mitigating unintended consequences of AI use. This applies, for instance, to the regulation of biases causing discrimination or to unsafe AI solutions threatening human life in areas such as health or autonomous driving. The search for global minimum standards constitutes a further prospective area of global AI regulation. Prospective candidates of international condemnation and prohibition could be subliminal AI techniques, manipulative AI solutions, or AI-powered lie detectors, but also the more general question of (biometric) data input. Furthermore, international regulation is relevant for AI use in warfare, but supervision technologies are also of critical importance for realizing existent codifications on human rights or the *jus in bello*.

Identifying Normative Boundaries

Normative limitations pertain to areas where AI cooperation or detrimental to *moral desiderata* in international affairs. The right to development and human rights in their social and economic dimensions correspond to the interest of the global majority in establishing global minimum standards. Here, decision makers should refrain from unilateral measures enhancing preexisting technological divides. Likewise, uncontrolled technology transfer can exacerbate preexisting human rights violations. Following the UN Guiding Principles (GPs), organizations ought not to be complicit in any human rights violations, implying that the proliferation of certain AI technologies to authoritarian regimes – such as AI-based lie detectors or AI solutions involving biometric data – would violate the UN GPs.

Involving Proportionality

Both implications are likely to collide in practice, especially when human rights-based restrictions on AI diffusion drive global inequalities. Hence, a balanced assessment requires minding the interest of individuals living in authoritarian regimes. As an approximation, we formulate the following premises as variables in the equation of limiting or enhancing technology transfer.

1. *Developing countries* (even the less democratic ones) require technological support in terms of AI solutions for closing preexisting gaps. Political concessions such as guarantees not to use AI solutions in specific domains could be realized within this collaborative format.
2. *Certain political structures* (with a low degree of checks and balances and low regard for human rights) and the existence of preexisting conflicts

present determine the likelihood of AI development and deployment at the expense of human rights and humanitarian law.

3. Specific AI solutions are *closely related to aims in international affairs*. This includes AI solutions to climate change (smart energy solutions) and those conducive to the resolution of a health crisis (AI-powered vaccine development). These AI solutions should be prioritized within global AI cooperation.
4. Areas of AI use that are generally deemed as *conflicting with human rights* include certain supervision technologies or the use of biometric data. Regulation on such technologies presents an urgent matter for international discourse.

These points might enable a cost/benefit assessment of AI cooperation from a moral perspective. However, international regulation and global diffusion of AI hinge on each other, as stalling international regulation might delay technology diffusion. Likewise, limiting technology transfer might hinder international alignment in a setting involving not only traditional powers such as the U.S., EU, and Russia, but also emerging leaders in particular fields of AI such as China, Turkey, and South Korea.

Conclusion

The question of how to engage in international cooperation on AI is an intricate issue. International AI cooperation itself is a complex topic involving two interdependent pillars, regulation and diffusion of AI. The preceding analysis indicates that norms are limiting but also guiding potential areas of AI cooperation. The configuration of AI as a basic innovation that might create a series of unintended consequences requires a higher degree of collaboration than other technologies. This requirement is exacerbated by interdependences and the urgency of global standards on critical AI solutions. AI diffusion is even more complex, due to the nature of AI as a powerful tool for development and its role in human rights violations. Policy makers need to include both aspects, representing different types of rights and interests in their calculations. For solving conflicts between both pillars, it is therefore important to distinguish not only between authoritarian and democratic nations, but also between developing and developed nations.

While the principle of proportionality presents a starting point for navigating international AI cooperation, more research needs to be undertaken to gain a more precise picture of further potential trade-offs in the global diffusion of AI.