



TECHNICAL REPORT

# Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development

Peter Cihon<sup>1</sup>

Research Affiliate, Center for the Governance of AI  
Future of Humanity Institute, University of Oxford  
[petercihon@gmail.com](mailto:petercihon@gmail.com)

*April 2019*

---

<sup>1</sup> For helpful comments on earlier versions of this paper, thank you to Jade Leung, Jeffrey Ding, Ben Garfinkel, Allan Dafoe, Matthijs Maas, Remco Zwetsloot, David Hageböling, Ryan Carey, Baobao Zhang, Cullen O’Keefe, Sophie-Charlotte Fischer, and Toby Shevlane. Thank you especially to Markus Anderljung, who helped make my ideas flow on the page. This work was funded by the Berkeley Existential Risk Initiative. All errors are mine alone.

## Executive Summary

Artificial Intelligence (AI) presents novel policy challenges that require coordinated global responses.<sup>2</sup> Standards, particularly those developed by existing international standards bodies, can support the global governance of AI development. International standards bodies have a track record of governing a range of socio-technical issues: they have spread cybersecurity practices to nearly 160 countries, they have seen firms around the world incur significant costs in order to improve their environmental sustainability, and they have developed safety standards used in numerous industries including autonomous vehicles and nuclear energy. These bodies have the institutional capacity to achieve expert consensus and then promulgate standards across the world. Other existing institutions can then enforce these nominally voluntary standards through both *de facto* and *de jure* methods.

AI standards work is ongoing at ISO and IEEE, two leading standards bodies. But these ongoing standards efforts primarily focus on standards to improve market efficiency and address ethical concerns, respectively. There remains a risk that these standards may fail to address further policy objectives, such as a culture of responsible deployment and use of safety specifications in fundamental research. Furthermore, leading AI research organizations that share concerns about such policy objectives are conspicuously absent from ongoing standardization efforts.

Standards will not achieve all AI policy goals, but they are a path towards effective global solutions where national rules may fall short. Standards can influence the development and deployment of particular AI systems through product specifications for, i.a., explainability, robustness, and fail-safe design. They can also affect the larger context in which AI is researched, developed, and deployed through process specifications.<sup>3</sup> The creation, dissemination, and enforcement of international standards can build trust among participating researchers, labs, and states. Standards can serve to globally disseminate best practices, as previously witnessed in cybersecurity, environmental sustainability, and quality management. Existing international treaties, national mandates, government procurement requirements, market incentives, and global harmonization pressures can contribute to the spread of standards once they are established. Standards do have limits, however: existing market forces are insufficient to incentivize the adoption of standards that govern fundamental research and other transaction-distant systems and practices. Concerted efforts among the AI community and external stakeholders will be needed to achieve such standards in practice.

---

<sup>2</sup> See, e.g., Brundage, Miles, et al. "The malicious use of artificial intelligence: Forecasting, prevention, and mitigation." *Future of Humanity Institute and the Centre for the Study of Existential Risk*. (2018) <https://arxiv.org/pdf/1802.07228.pdf>; Dafoe, Allan. AI Governance: A Research Agenda. *Future of Humanity Institute*. (2018). [www.fhi.ox.ac.uk/wp-content/uploads/GovAI/Agenda.pdf](http://www.fhi.ox.ac.uk/wp-content/uploads/GovAI/Agenda.pdf); Bostrom, Nick, Dafoe, Allan and Carrick Flynn. "Public Policy and Superintelligent AI: A Vector Field Approach." (working paper, Future of Humanity Institute, 2018), <https://nickbostrom.com/papers/aipolicy.pdf>; Cave, Stephen, and Seán S. ÓhÉigartaigh. "Bridging near-and long-term concerns about AI." *Nature Machine Intelligence* 1, no. 1 (2019): 5.

<sup>3</sup> For discussion of the importance of context in understanding risks from AI, see Zwetsloot, Remco and Allan Dafoe. "Thinking About Risks From AI: Accidents, Misuse and Structure." *Lawfare*. (2019). <https://www.lawfareblog.com/thinking-about-risks-ai-accidents-misuse-and-structure>.

Ultimately, standards are a tool for global governance, but one that requires institutional entrepreneurs to actively use standards in order to promote beneficial outcomes. Key governments, including China and the U.S., have stated priorities for developing international AI standards. Standardization efforts are only beginning, and may become increasingly contentious over time, as has been witnessed in telecommunications. Engagement sooner rather than later can establish beneficial and internationally legitimate ground rules to reduce risks in international and market competition for the development of increasingly capable AI systems.

In light of the strengths and limitations of standards, this paper offers a series of recommendations. They are summarized below:

- **Leading AI labs should build institutional capacity to understand and engage in standardization processes.** This can be accomplished through in-house development or partnerships with specific third-party organizations.
- **AI researchers should engage in ongoing standardization processes.** The Partnership on AI and other qualifying organizations should consider becoming liaisons with standards committees to contribute to and track developments. Particular standards may benefit from independent development initially and then be transferred to an international standards body under existing procedures.
- **Further research is needed on AI standards from both technical and institutional perspectives.** Technical standards desiderata can inform new standardization efforts and institutional strategies can develop paths for standards spread globally in practice.
- **Standards should be used as a tool to spread a culture of safety and responsibility among AI developers.** This can be achieved both inside individual organizations and within the broader AI community.

# Table of Contents

<b>Executive Summary</b>	<b>2</b>
<b>Table of Contents</b>	<b>4</b>
<b>Glossary</b>	<b>5</b>
<b>1. Introduction</b>	<b>6</b>
<b>2. Standards: Institution for Global Governance</b>	<b>7</b>
2A. The need for global governance of AI development	7
2B. International standards bodies relevant to AI	9
2C. Advantages of international standards as global governance tools	10
2Ci. Standards Govern Technical Systems and Social Impact	11
2Cii. Shaping expert consensus	14
2Ciii. Global reach and enforcement	15
<b>3. Current Landscape for AI Standards</b>	<b>19</b>
3A. International developments	19
3B. National priorities	21
3C. Private initiatives	24
<b>4. Recommendations</b>	<b>25</b>
4A. Engage in ongoing processes	25
4Ai. Build capacity for effective engagement	26
4Aii. Engage directly in ongoing processes	27
4Aiii. Multinational organizations should become liaisons	28
4B. Pursue parallel standards development	28
4C. Research standards and strategy for development	29
4Ci. Research technical standards desiderata	29
4Cii. Research strategies for standards in global governance	30
4D. Use standards as a tool for culture change	31
<b>5. Conclusion</b>	<b>32</b>
<b>References</b>	<b>33</b>
<b>Appendix 1: ISO/IEC JTC 1 SC 42 Ongoing Work</b>	<b>40</b>
<b>Appendix 2: IEEE AI Standards Ongoing Work</b>	<b>41</b>

# Glossary

<b>Acronym</b>	<b>Meaning</b>
DoD	U.S. Department of Defense
CFIUS	Committee on Foreign Investment in the United States
ECPAIS	IEEE Ethics Certification Program for Autonomous and Intelligent Systems
GDPR	EU General Data Protection Regulation
IEEE	Institute of Electrical and Electronics Engineers
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
ITU	International Telecommunications Union
JTC 1	Joint Technical Committee 1, formed by IEC and ISO to create information technology standards
MNC	Multinational corporation
OCEANIS	Open Community for Ethics in Autonomous and Intelligent Systems
TBT	WTO Agreement on Technical Barriers to Trade
WTO	World Trade Organization

# 1. Introduction

Standards are an institution for coordination. Standards ensure that products made around the world are interoperable. They ensure that management processes for cybersecurity, quality assurance, environmental sustainability, and more are consistent no matter where they happen. Standards provide the institutional infrastructure needed to develop new technologies, and they provide safety procedures to do so in a controlled manner. Standards can do all of this, too, in the research and development of artificial intelligence (AI).

Market incentives will drive companies to participate in the development of product standards for AI. Indeed, work is already underway on preliminary product and ethics standards for AI. But, absent outside intervention, standards may not serve as a policy tool to reduce risks in the technology's development.<sup>4</sup> Leading AI research organizations that share concerns about such risks are conspicuously absent from ongoing standardization efforts.<sup>5</sup> To positively influence the development trajectory of AI, we do not necessarily need to design new institutions. Existing organizations, treaties, and practices already see standards disseminated around the world, enforced through private institutions, and mandated by national action.

Standards, developed by an international group of experts, can provide legitimate global rules amid international competition in the development of advanced AI systems.<sup>6</sup> These standards can support trust among developers and a consistent focus on safety, among other benefits. Standards constitute a language and practice of communication among research labs around the world, and can establish guardrails that help support positive AI research and development outcomes.

Standards will not achieve all AI policy goals, but they are an important step towards effective global solutions. They are an important step that the AI research community can start leading on today. The paper is structured as follows. *Section 2* discusses the need for global coordination on AI policy goals and develops at length the use of international standards in achieving these goals. *Section 3* analyzes the current AI standards landscape. *Section 4* offers a series of recommendations for how the AI community, comprising technical researchers, development organizations, governance researchers, can best use international standards as a tool of global governance.

---

<sup>4</sup> This work fits within a growing literature that argues that short-term and long-term AI policy should not be considered separately. Policy decisions today can have long-term implications. See, e.g., Cave and ÓhÉigeartaigh, "Bridging near-and long-term concerns about AI."

<sup>5</sup> Some in the AI research community do acknowledge the significance of standards, but they see efforts towards standardization as a future endeavor: the [OpenAI Charter](#) acknowledges the importance of sharing standards research, but focused on a time when they curtail open publication. The [Partnership on AI](#) is today committed to establishing best practices on AI, in contrast to formal standards.

<sup>6</sup> Advanced AI incorporates future developments in machine intelligence substantially more capable than today's systems but at a level well short of an Artificial General Intelligence. See Dafoe, "AI Governance: A Research Agenda."

## 2. Standards: Institution for Global Governance

### 2A. The need for global governance of AI development

AI development poses global challenges. Government strategies to incentivize increased AI research within national boundaries may result in a fractured governance landscape globally,<sup>7</sup> and in the long-term threaten a race to the bottom in regulatory stringency.<sup>8</sup> In this scenario, countries compete to attract AI industry through national strategies and incentives that accelerate AI development, but do not similarly increase regulatory oversight to mitigate societal risks associated with these developments.<sup>9</sup> These risks associated with lax regulatory oversight and heated competition range from increasing the probability of biased, socially harmful systems<sup>10</sup> to existential threats to human life.<sup>11</sup>

These risks are exacerbated by a lack of effective global governance mechanisms to provide, at minimum, guardrails in the competition that drives technological innovation. Although there is uncertainty surrounding AI capability development timelines,<sup>12</sup> AI researchers expect capabilities to match human performance for many tasks within the decade and for most tasks within several decades.<sup>13</sup> These developments will have transformative effects on society.<sup>14</sup> It is thus critical that global governance institutions are put in place to steer these transformations in beneficial directions.

International standards are an institution of global governance that exists today and can help achieve AI policy goals. Notably, global governance does not mean global government: existing regimes of international coordination, transnational collaboration, and global trade are all forms of global governance.<sup>15</sup> Not all policy responses to AI will be global; indeed, many will necessarily account for local and national contexts.<sup>16</sup> But

---

<sup>7</sup> Cihon, Peter. "Regulatory Dynamics of Artificial Intelligence Global Governance." *Typhoon Consulting*. (2018). <http://www.typhoonconsulting.com/wp-content/uploads/2018/07/18.07.11-AI-Global-Governance-Peter-Cihon.pdf>.

<sup>8</sup> See Dafoe, "AI Governance: A Research Agenda."

<sup>9</sup> Indeed, labs pursuing AI systems with advanced capabilities are globally distributed and demonstrate a high variance in their operating and safety procedures. Baum, Seth. "A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy." *Global Catastrophic Risk Institute Working Paper 17-1*. (2017).

<sup>10</sup> Whittaker, Meredith, Kate Crawford, Roel Dobbe, et al. "AI Now Report 2018." *AI Now*. (2018). [https://ainowinstitute.org/AI\\_Now\\_2018\\_Report.pdf](https://ainowinstitute.org/AI_Now_2018_Report.pdf).

<sup>11</sup> Bostrom, Nick. *Superintelligence*. Oxford: Oxford University Press, 2014.

<sup>12</sup> Past technologies have seen discontinuous progress, but this may not come to pass in development of advanced AI. See blog posts on AI Impacts from [2015](#) and [2018](#).

<sup>13</sup> Grace, Katja, John Salvatier, Allan Dafoe, Baobao Zhang, and Owain Evans. "When will AI exceed human performance? Evidence from AI experts." *Journal of Artificial Intelligence Research* 62 (2018): 729-754; See Bughin, Jacques, Jeongmin Seong, James Manyika, et al. "Notes from the AI Frontier: Modeling the Impact of AI on the World Economy." *McKinsey Global Institute*. (2018). <https://www.mckinsey.com/~media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.ashx>.

<sup>14</sup> Dafoe, "AI Governance: A Research Agenda"; Bostrom, Dafoe, Flynn. "Public Policy and Superintelligent AI."

<sup>15</sup> Hägel, Peter. "Global Governance." *International Relations*. Oxford: Oxford University Press, 2011.

<sup>16</sup> See, e.g. Awad, Edmond, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. "The moral machine experiment." *Nature* 563, no. 7729 (2018): 59.

international standards can support policy goals where global governance is needed, in particular, by (1) spreading beneficial systems and practices, (2) facilitating trust among states and researchers, and (3) encouraging efficient development of advanced systems.<sup>17</sup>

First, the content of the standards themselves can support AI policy goals. Beneficial standards include those that support the security and robustness of AI, further the explainability of and reduce bias in algorithmic decisions, and ensure that AI systems fail safely. Standards development on all three fronts is underway today, as discussed below in *Section 3*. Each standard could also reduce long-term risks if their adoption shifts funding away from opaque, insecure, and unsafe methods.<sup>18</sup> Additional standards could shape processes of research and development towards beneficial ends, namely through an emphasis on safety practices in fundamental research. In addition to stipulating safe processes, these standards, through their regular enactment and enforcement, could encourage a responsible culture of AI development. These claims are developed further in *Section 4*.

Second, international standards processes can facilitate trust among states and research efforts. International standards bodies provide focal organizations where opposing perspectives can be reconciled. Once created and adopted, international standards can foster trust among possible competitors because they will provide a shared governance framework from which to build further agreement.<sup>19</sup> This manner of initial definition, measurement, or other initial agreement contributing to subsequent expanded and enforced agreements has been witnessed in other international coordination problems, e.g., nuclear test ban treaties and environmental protection efforts.<sup>20</sup> Trust is also dependent on the degree of open communication among labs. Complete openness can present problems; indeed, open publication of advanced systems,<sup>21</sup> and even simply open reporting of current capabilities in the future could present significant risks.<sup>22</sup> Standards can facilitate *partial* openness among research efforts that is “unambiguously good” in light of these concerns.<sup>23</sup> In practice, credible public commitments to specific standards can provide partial information about the practices of otherwise disconnected labs. Furthermore, particular standards that may emerge over time could themselves define appropriate levels and mechanisms of openness.

Third, international standards can encourage the efficient development of increasingly advanced AI systems. International standards have a demonstrated track record of improving global market efficiency and economic surplus via, i.a., reduced barriers to international trade, greater interoperability of labor and end-products, and eliminated duplicated effort on standardized elements.<sup>24</sup> International standards could support these outcomes

---

<sup>17</sup> These are key policy elements that bridge a focus on current systems with a long-term research towards superintelligent AI. See Bostrom, Dafoe, Flynn. “Public Policy and Superintelligent AI.”

<sup>18</sup> See Cave and ÓhÉigeartaigh, “Bridging near-and long-term concerns about AI.”

<sup>19</sup> Bostrom, Nick. “Strategic implications of openness in AI development.” *Global Policy* 8, no. 2 (2017): 146.

<sup>20</sup> For example, the Vienna Convention for the Protection of the Ozone Layer provided a framework for a later agreement in the Montreal Protocol that has seen global adoption and enforcement.

<sup>21</sup> See, e.g., OpenAI’s limited release of its [GPT-2](#) natural language model.

<sup>22</sup> Bostrom, “Strategic implications of openness in AI development.”

<sup>23</sup> *Ibid.*, 145.

<sup>24</sup> See Büthe, Tim, and Walter Mattli. *The new global rulers: The privatization of regulation in the world economy*. Princeton University Press, 2011; Brunsson, Nils and Bengt Jacobsson. “The pros and cons of standardization—an epilogue” in *A World of Standards*. Oxford: Oxford University Press, 2000, 169-173; Abbott, Kenneth W., and Duncan Snidal. “International ‘standards’ and international governance.” *Journal of European Public Policy* 8, no. 3 (2001): 345.



for AI as well, e.g., with systems that can deploy across national boundaries and be implemented using consistent processes and packages by semi-skilled AI practitioners. Increased efficiency in deployment will drive further resources into research and development. Some in the AI community may be concerned that this will increase the rate at which AI research progresses, thereby encouraging racing dynamics that disincentivize precaution.<sup>25</sup> Yet standards can help here too, both through object-level standards for safety practices with enforcement mechanisms and by facilitating trust among developers. These claims are developed further in *Sections 2C* and *4*.

In summary, continued AI development presents risks that require coordinated global governance responses. International standards are an existing form of global governance that can offer solutions. These standards can help support efficient development of AI industry, foster trust among states and developers of the technology, and see beneficial systems and practices enacted globally. It is important to note that, regardless of intervention, ongoing standards work will encourage increased efficiency in AI development. Engagement is needed to support standards that help foster trust and encourage beneficial systems and processes globally.

## 2B. International standards bodies relevant to AI

A wide range of organizations develop standards that are adopted around the world. AI researchers may be most familiar with proprietary or open-source software standards developed by corporate sponsors, industry consortia, and individual contributors. These are common in digital technologies, including the development of AI, e.g., software libraries including TensorFlow, PyTorch, and OpenAI Gym that become standards across industry over time.<sup>26</sup> The groups responsible for such standards, however, do not have experience in monitoring and enforcement of such standards globally. In contrast, international standards bodies have such experience. This section discusses standards bodies, and *Section 2C* describes relevant categories of standards.

Specialized bodies may create international standards. These bodies can be treaty organizations such as the International Atomic Energy Agency and the International Civil Aviation Organization, which govern standards on nuclear safety and international air travel, respectively. Such a body may well suit the governance of AI research and development, but its design and implementation are beyond the scope of this paper. Instead, this paper focuses on *existing* institutions that can host development of needed standards and see them enacted globally. Non-state actors' efforts towards institutional development tend to be more successful in both agenda setting and impact if they work with states and seek change that can be accommodated within existing structures and organizations.<sup>27</sup> Thus, existing international standards bodies present an advantage. Nevertheless,

---

<sup>25</sup> Armstrong, S., Bostrom, N. & Shulman, C. "Racing to the precipice: a model of artificial intelligence development", Technical Report #2013-1, Future of Humanity Institute, Oxford University: pp. 1-8. (2013). <https://www.fhi.ox.ac.uk/wp-content/uploads/Racing-to-the-precipice-a-model-of-artificial-intelligence-development.pdf>.

<sup>26</sup> Software libraries, programming languages, and operating systems are standards insofar as they guide behavior. They may not emerge from standardization processes but instead market competition. See *Section 3C*.

<sup>27</sup> Hale, Thomas, and David Held. *Beyond Gridlock*. Cambridge: Polity. 2017.

if a specialized agency is developed in the future, previously established standards can be incorporated at that time.<sup>28</sup>

There are two existing international standards bodies that are currently developing AI standards. First is a joint effort between ISO and IEC. To coordinate development of digital technology standards, ISO and IEC established a joint committee (JTC 1) in 1987. JTC 1 has published some 3000 standards, addressing everything from programming languages, character renderings, file formats including JPEG, distributed computing architecture, and data security procedures.<sup>29</sup> These standards have influence and have seen adoption and publicity by leading multinational corporations (MNCs). For example, ISO data security standards have been widely adopted by cloud computing providers, e.g., Alibaba, Amazon, Apple, Google, Microsoft, and Tencent.

30

The second international standards body that is notable in developing AI standards is the IEEE Standards Association. IEEE is an engineers' professional organization with a subsidiary Standards Association (SA) whose most notable standards address protocols for products, including Ethernet and WiFi. IEEE SA also creates process standards in other areas including software engineering management and autonomous systems design. Its AI standardization processes are part of a larger IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems.<sup>31</sup>

A third international standards body may become increasingly relevant for AI in the future: the ITU. The ITU has historically played a role in standards for information and communications technologies, particularly in telecommunications. It has a Focus Group on Machine Learning for Future Networks that falls within this telecommunications remit. Following the 2018 AI for Good Global Summit, it has also created a Focus Group on AI for Health, "which aims *inter alia* to create standardized benchmarks to evaluate Artificial Intelligence algorithms used in healthcare applications."<sup>32</sup> Given the ITU's historically narrower scope, however, this paper does not consider the organizations' work further.

## 2C. Advantages of international standards as global governance tools

International standards present a number of advantages in encouraging the global governance of AI. This section distills these advantages into three themes. First, international standards have a history of guiding the development and deployment of technical systems and shaping their social effects across the world. Second, international standards bodies privilege the influence of experts and have tested mechanisms for achieving

---

<sup>28</sup> Pre-existing standards have been referenced in international treaties, e.g., The International Maritime Organization's Safety of Life at Sea Treaty references ISO product standards. Koppell, Jonathan G. S. *World Rule: Accountability, Legitimacy, and the Design of Global Governance*. Chicago: University of Chicago Press, 2010.

<sup>29</sup> See, e.g., Rajchel, Lisa. *25 years of ISO/IEC JTC 1*. ISO Focus+, 2012.

[https://www.iso.org/files/live/sites/isoorg/files/news/magazine/ISO%20Focus%2b%20\(2010-2013\)/en/2012/ISO%20Focus%2b%2c%20June%202012.pdf](https://www.iso.org/files/live/sites/isoorg/files/news/magazine/ISO%20Focus%2b%20(2010-2013)/en/2012/ISO%20Focus%2b%2c%20June%202012.pdf).

<sup>30</sup> For Amazon, these include ISO 27001, 27017, and 27018 from JTC 1 as well as the ISO 9001 quality management process standard. See the link associated with each company: [Alibaba](#), [Amazon](#), [Apple](#), [Google](#), [Microsoft](#), and [Tencent](#).

<sup>31</sup> See [IEEE's Ethics in Action website](#).

<sup>32</sup> See [ITU's AI for Good website](#).

consensus among them on precisely what should be in standards. Third, existing treaties, national practices, and transnational actors encourage the global dissemination and enforcement of international standards.

## 2Ci. Standards Govern Technical Systems and Social Impact

Standards are, at their most fundamental “a guide for behavior and for judging behavior.”<sup>33</sup> In practice, standards define technical systems and can guide their social impact. Standards are widely used for both private and public governance at national and transnational levels, in areas as wide ranging as financial accounting and nuclear safety.<sup>34</sup> Many forms of standards will impact the development of AI.

Consider a useful typology of standards based on actors’ incentives and the object of standardization. Actors’ incentives in standards can be modeled by two types of externalities: positive, network externalities and negative externalities.<sup>35</sup> With network externalities, parties face a coordination game where they are incentivized to cooperate.<sup>36</sup> For example, a phone is more useful if it can call many others than if it can only communicate with the same model. Institutions may be necessary to establish a standard in this case but not to maintain the standard in practice, as the harmony of interests obviates enforcement. For the purposes of this paper, consider these standards “network standards.”

Negative externalities are different; a polluter burdens others but does not internalize the cost itself. Standards here face challenges: individuals may have an incentive to defect in what could be modeled as a Prisoner’s Dilemma.<sup>37</sup> In the pollution case, it is in the interest of an individual business to disregard a pollution standard absent additional institutions. But this interest can favor cooperation if an institution creates excludible benefits and an enforcement mechanism. External stakeholders are important here as well: institutions to enable enforced standards are incentivized by demand external to those who adopt the standards. In practice, governments, companies, and even public pressure can offer such incentives; many are explored in *Section 2Ciii*. For example, the ISO 14001 Environmental Management standard requires regular and intensive audits in order to obtain certification, which in turn brings reputational value to the companies that obtain it.<sup>38</sup> In general, for such standards, institutions are needed for initial standardization *and* subsequent enforcement. For the purposes of this paper, consider these standards “enforced standards.” Enforcement can take multiple forms, from regulatory mandates to contractual monitoring. Certification of adherence to a standard is a common method of enforcement that relies on third-parties, which can be part of government or private entities.<sup>39</sup> Self-certification is also common, whereby a firm will claim that it complies with a standard and is subject to

---

<sup>33</sup> Abbot and Snidal, “International ‘standards’ and international governance,” p. 345.

<sup>34</sup> Brunsson, Nils and Bengt Jacobsson. “The contemporary expansion of standardization” in *A World of Standards*. Oxford: Oxford University Press, 2000, 1-18.

<sup>35</sup> Abbott and Snidal, “International ‘standards’ and international governance.”

<sup>36</sup> This scenario can have distributional consequences as well, where one party gains more from the standard, but ultimately all are better off from cooperation.

<sup>37</sup> Abbot and Snidal, “International ‘standards’ and international governance.”

<sup>38</sup> Prakash, Aseem, and Matthew Potoski. *The voluntary environmentalists: Green clubs, ISO 14001, and voluntary environmental regulations*. Cambridge University Press, 2006.

<sup>39</sup> Ibid.

future enforcement from a regulator.<sup>40</sup> Compliance monitoring can occur through periodic audits, applications for re-certification, or ad hoc investigations in response to a whistleblower or documented failure.<sup>41</sup> In summary, both categories of standards exist--network and enforced--but enforced standards require additional institutions for successful implementation.

In practice, standards address one of two objects: products or management processes. Product standards can define terminology, measurements, variants, functional requirements, qualitative properties, testing methods, and labeling criteria.<sup>42</sup> Management process standards can describe processes or elements of organizations to achieve explicit goals, e.g., quality, sustainability, and software life cycle management. A process that follows a particular standard need not impose costs with each iteration of a product: the standardized process simply informs how new products are created. Indeed, process standards can often function as a way for firms to adopt best practices in order to increase their competitiveness.<sup>43</sup> One such ISO standard on cybersecurity has been adopted by firms in nearly 160 countries.<sup>44</sup> Figure 1 illustrates these standards categories as they relate to externalities with some notable examples.

Standards for AI will emerge in all four quadrants; indeed, as discussed below in *Section 3*, standards that span the typology are already under development. Different types of standards will spread with more or less external effort, however. Network-product standards that support interoperability and network-process standards that offer best practices will see actors adopt them in efforts to grow the size of their market and reduce their costs. Indeed, most international standards from ISO/IEC and IEEE are product standards that address network externalities, seeking to increase the interoperability of global supply chains.<sup>45</sup> Enforced standards will require further incentivization from external stakeholders, whether they be regulators, contracting companies, or the public at large. The more distant the object of standardization is from common market transactions, the more difficult the incentivization of standards will be without external intervention. In particular, this means that an enforced-process standard for safety in basic research and development is unlikely to develop without concerted effort from the AI community.

---

<sup>40</sup> Firms may declare that their practices or products conform to network standards, even, in some cases, choosing to certify this conformity. In these cases, however, the certification serves as a signal to access network benefits. Although enforced standards are not the only category that may see certification, it is the category that requires further enforcement to address possible incentives to defect.

<sup>41</sup> Some AI standards, namely those on safety of advanced research, will benefit from novel monitoring regimes. See *Section 4*.

<sup>42</sup> Hallström, Kristina Tamm. *Organizing International Standardization: ISO and the IASC in Quest of Authority*. Cheltenham: Edward Elgar, 2004.

<sup>43</sup> Brunsson and Jacobsson. "The pros and cons of standardization."

<sup>44</sup> "The ISO Survey of Management System Standard Certifications - 2017 - Explanatory Note. ISO. Published August, 2018.

[https://isotc.iso.org/livelink/livelink/fetch/-8853493/8853511/8853520/18808772/00.Overall\\_results\\_and\\_explanatory\\_note\\_on\\_2017\\_Survey\\_results.pdf?nodeid=19208898&vernum=-2](https://isotc.iso.org/livelink/livelink/fetch/-8853493/8853511/8853520/18808772/00.Overall_results_and_explanatory_note_on_2017_Survey_results.pdf?nodeid=19208898&vernum=-2); ISO 27001 had nearly 40,000 certifications in 159 countries in 2017. See [ISO 27001 website](#).

<sup>45</sup> Bütthe and Mattli, *The new global rulers*.

**Figure 1. Standards typology with examples**

<b>Network - Product</b>	<b>Network - Process</b>
<p>Protocols for establishing Wi-Fi connections (IEEE 802.11)</p> <p>Standard dimensions for a shipping container to enable global interoperability (ISO 668)</p>	<p>Quality management process standard that facilitates international contracting and supply chains by ensuring consistency globally (ISO 9001)</p> <p>Information security management system, requirements and code of practice for implementation and maintenance. (ISO/IEC 27001 and 27002, respectively)</p> <p>Software life cycle management processes (ISO/IEC/IEEE 12207)</p>
<b>Enforced - Product</b>	<b>Enforced - Process</b>
<p>Paper products sourced with sustainable methods and monitored through supply chain (Forest Stewardship Council)<sup>46</sup> <i>Enforced:</i> Third-party certified.</p> <p>CE marking for safety, health, and environmental protection requirements for sale within the European Economic Area. (EU) <i>Enforced:</i> If problems arise, violations are sanctioned by national regulators.<sup>47</sup></p>	<p>Environmental Management process standard helps organizations minimize the environmental footprint of their operations (ISO 14001) <i>Enforced:</i> Third-party certified.</p> <p>Functional safety management over life cycle for road vehicles (ISO 26262). <i>Enforced:</i> Required to meet safety regulations and import criteria.</p> <p>Safety requirements for collaborative industrial robots (ISO/TS 15006). <i>Enforced:</i> Supports obligations under safety regulations.</p>

There are, however, also notable examples of enforced standards that do see firms take on considerable costs to internalize harmful externalities. The ISO 14001 Environmental Management standard has spread to 171 countries, and saw over 360,000 certifications around the world in 2017.<sup>48</sup> This standard provides firms a framework to improve the environmental sustainability of their practices and certification demonstrates that they have done so in order to gain reputational benefits from environmental regulators.<sup>49</sup> Firms take on significant costs in certification, the total process for which can cost upwards of \$100,000 per facility.<sup>50</sup> The standard has been notable for spreading sustainable practices to middle-tier firms that do not differentiate

<sup>46</sup> See the certification description on the [Forest Stewardship Council website](#).

<sup>47</sup> See [European Commission website on CE marking](#).

<sup>48</sup> “The ISO Survey of Management System Standard Certifications - 2017 - Explanatory Note. *ISO*. Published August, 2018.

[https://isotc.iso.org/livelink/livelink/fetch/-8853493/8853511/8853520/18808772/00\\_Overall\\_results\\_and\\_explanatory\\_note\\_on\\_2017\\_Survey\\_results.pdf?nodeid=19208898&vernum=-2](https://isotc.iso.org/livelink/livelink/fetch/-8853493/8853511/8853520/18808772/00_Overall_results_and_explanatory_note_on_2017_Survey_results.pdf?nodeid=19208898&vernum=-2); ISO 14001 had over 360,000 certifications in 171 countries in 2017. See [ISO 14000 website](#).

<sup>49</sup> Prakash, Aseem, and Matthew Potoski. *The voluntary environmentalists*.

<sup>50</sup> Ibid.

themselves based on environmentally sustainable practices.<sup>51</sup> Clearly, however, ISO 14001 has not solved larger environmental challenges. The narrower success of this program should inform expectations of the role for standards for AI development: although they can encourage global adoption of best practices and see firms incur significant costs to undertake them, standards will not be a complete solution.

Another category of enforced standards relevant to AI standards are product and process safety standards. Safety standards for medical equipment, biological lab processes, and safety in human-robot collaboration have been spread globally by international standards bodies and other international institutions.<sup>52</sup> Related standards for functional safety, i.e., processes to assess risks in operations and reduce them to tolerable thresholds, are widely used across industry, from autonomous vehicle development to regulatory requirements for nuclear reactor software.<sup>53</sup> These standards do not apply directly to the *process* of cutting-edge research. That is not to say, however, that with concerted effort new standards guided by these past examples could not do so.

## 2Cii. Shaping expert consensus

The internal processes of international standards bodies share two characteristics that make them useful for navigating AI policy questions. First, these bodies privilege expertise. Standards themselves are seen as legitimate rules to be followed precisely because they reflect expert opinion.<sup>54</sup> International standards bodies generally require that any intervention to influence a standard must be based in technical reasoning.<sup>55</sup>

This institutional emphasis on experts can see an individual researcher's engagement be quite impactful. Unlike other methods of global governance that may prioritize experts, e.g., UN Groups of Governmental Experts which yield mere advice, experts involved in standards organizations have influence over standards that can have *de facto* or even *de jure* governing influence globally. Other modes of *de jure* governance, e.g., national

---

<sup>51</sup> Ibid.

<sup>52</sup> There is no data available on uptake of ISO/TS 15006 Robots and robotic devices -- Collaborative robots or related standards. ISO does claim, however, that IEC 60601 and ISO 10993 have seen global recognition and uptake for ensuring safety in medical equipment and biological processes, respectively. See discussion of [ISO standards in sectorial examples at ISO's dedicated webpage](#).

<sup>53</sup> Particular industry applications derive from the generic framework standard for functional safety, IEC 61508, including ISO 26262 Road vehicles -- Functional safety and IEC 61513 Nuclear power plants - Instrumentation and control for systems important to safety - General requirements for systems. Per conversation with an employee at a firm developing autonomous driving technology, all teams in the firm have safety strategies that cite the standard. Nuclear regulators reference the relevant standard, see, e.g., IAEA. *Implementing Digital Instrumentation and Control Systems in the Modernization of Nuclear Power Plants*. Vienna: IAEA. 2009. [https://www-pub.iaea.org/MTCD/Publications/PDF/Pub1383\\_web.pdf](https://www-pub.iaea.org/MTCD/Publications/PDF/Pub1383_web.pdf).

See generally, Smith, David J., and Kenneth GL Simpson. *Safety critical systems handbook: a straight forward guide to functional safety, IEC 61508 (2010 Edition) and related standards, including process IEC 61511 and machinery IEC 62061 and ISO 13849*. Elsevier, 2010.

<sup>54</sup> Murphy, Craig N., and JoAnne Yates. *The International Organization for Standardization (ISO): global governance through voluntary consensus*. London: Routledge, 2009; Jacobsson, Bengt. "Standardization and expert knowledge" in *A World of Standards*. Oxford: Oxford University Press, 2000, 40-50.

<sup>55</sup> This is not to say that standards are apolitical. Arguments made with technical reasoning do not realize a single, objective standard; rather, technical reasoning can manifest in multiple forms of a particular standard, each with distributional consequences. See Büthe and Mattli, *The new global rulers*.

regulation or legislation, present only limited direct opportunities for expert engagement.<sup>56</sup> Some are concerned that such public engagement may undermine policy efforts on specific topics like AI safety.<sup>57</sup> Thus, for an AI researcher looking to maximize her global regulatory impact, international standards bodies offer an efficient venue for engagement.<sup>58</sup> Similarly, AI research organizations that wish to privilege expert governance may find international standards bodies a venue that has greater reach and legitimacy than closed self-regulatory efforts.

Second, standards bodies and their processes are designed to facilitate the arrival of consensus on what should and should not be within a standard.<sup>59</sup> This consensus-achieving experience is useful when addressing questions surrounding emerging technologies like AI that may face initial disagreement.<sup>60</sup> Although achieving consensus can take time, it is important to note that the definition of consensus in these organizations does not imply unanimity, and in practice it can often be achieved through small changes to facilitate compromise.<sup>61</sup> This institutional capacity to resolve expert disagreements based on technical argument stands in contrast to legislation or regulation that will impose an approach after accounting for limited expert testimony or filings. The capacity to resolve expert disagreement is important for AI, where it will help resolve otherwise controversial questions of what AI research is mature enough to include in standards.

### 2Ciii. Global reach and enforcement

International trade rules, national policies, and corporate strategy disseminate international standards globally. These mechanisms encourage or even mandate adoption of what are nominally voluntary standards. This section briefly describes these mechanisms and the categories of standards to which they apply. Taken together, these mechanisms can lead to the global dissemination and enforcement of AI standards.

International trade agreements are key levers for the global dissemination of standards. The World Trade Organization's Agreement on Technical Barriers to Trade (TBT) mandates that WTO member states use

---

<sup>56</sup> Congressional or parliamentary testimony does not necessarily translate into law and legislative staffers rarely have narrow expertise. Such limitations inform calls for a specialized AI regulatory agency within the US context. A regulatory agency can privilege experts, but only insofar as they work within the agency and, for example, forgo research. Standards organizations, alternatively, allow experts to continue their research work. On limitations of expertise in domestic governance and a proposal for an AI agency, see Scherer, Matthew U. "Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies." *Harv. JL & Tech.* 29 (2015): 353-400.

<sup>57</sup> See, e.g., Larks. "2018 AI Alignment Literature Review and Charity Comparison." [AI Alignment Forum blog post.](#)

<sup>58</sup> *Supra* footnote 55, expert participation remains political. Some standards are more politicized than others, though this does not follow a clear division between product or process, network or enforced. Vogel sees civil regulation (essentially enforced standards) as more politicized than technical (network) standards, although he looks at more politicized venues than simply international standards bodies. Network standards with large distributional consequences are often politicized, including shipping containers and ongoing 5G efforts. Vogel, David. "The Private Regulation of Global Corporate Conduct." in Mattli, Walter., and Ngaire. Woods, eds. *The Politics of Global Regulation*. Princeton: Princeton University Press, 2009; Bütte and Mattli, *The new global rulers*.

<sup>59</sup> See, e.g., Bütte and Mattli, *The new global rulers*, Chapter 6.

<sup>60</sup> Questions of safe procedures for advanced AI research, for instance, have not yet seen debate oriented towards consensus.

<sup>61</sup> Bütte and Mattli, *The new global rulers*, pp. 130-1.



international standards<sup>62</sup> where they exist, are effective, and are appropriate. This use can take two forms: incorporation into enforced technical regulations or into voluntary standards at the national level.<sup>63</sup> The TBT applies to existing regulations regardless if they are new or old; thus, if a new international standard is established, pre-existing laws can be challenged.<sup>64</sup> The TBT has a formal notice requirement for such regulation and enables member states to launch disputes within the WTO.<sup>65</sup>

There are important limitations to TBT, however. Few TBT-related disputes have been successfully resolved in the past.<sup>66</sup> TBT applies only to product and product-related process standards,<sup>67</sup> thus precluding its use in spreading standards on fundamental AI research. In a further limitation, the agreement permits national regulations to deviate from international standards in cases where “urgent problems of safety, health, environmental protection or national security arise,” although such cases require immediate notification and justification to the WTO.<sup>68</sup>

National policies are another key lever in disseminating international standards. National regulations reference international standards and can mandate compliance *de jure* in developed and developing countries alike.<sup>69</sup> Governments may use their purchasing power to encourage standards adoption via procurement requirements. EU member state procurement must draw on European or international standards where they exist,<sup>70</sup> and the optional WTO Agreement on Government Procurement encourages parties to use international standards for

---

<sup>62</sup> The TBT does not define international standards bodies, but it does set out a Code of Good Practice for standards bodies to follow and issue notifications of adherence to ISO. IEEE declared adherence to the Principles in 2017. See [ISO's WTO information gateway webpage](#); “IEEE Position Statement: IEEE Adherence to the World Trade Organization Principles for International Standardization.” *IEEE*. Published May 22, 2017. <http://globalpolicy.ieee.org/wp-content/uploads/2017/05/IEEE16029.pdf>.

<sup>63</sup> TBT does not mandate nations impose regulation; rather, it mandates that if they do so, regulation should incorporate international standards. Thus, nations may choose to simply leave a particular industry unregulated. This is unlikely, however, given international market pressures outline below.

<sup>64</sup> It does not, however, compel a national government to regulate in the first place. See Mattli, Walter. “The politics and economics of international institutional standards setting: an introduction.” *Journal of European Public Policy* 8, no. 3 (2001): 328-344.

<sup>65</sup> In 2015, there were some 25,000 notices of national regulatory measures, 473 concerns raised, and 6 disputes brought to the WTO. *Technical Barriers to Trade: Reducing trade friction from standards and regulations*. Geneva: WTO, 2015. [https://www.wto.org/english/thewto\\_e/20y\\_e/tbt\\_brochure2015\\_e.pdf](https://www.wto.org/english/thewto_e/20y_e/tbt_brochure2015_e.pdf).

<sup>66</sup> See Wijkström, Erik, and Devin McDaniels. “Improving Regulatory Governance: International Standards and the WTO TBT Agreement.” *Journal of World Trade* 47, no. 5 (2013): 1013-046.

<sup>67</sup> Scholars disagree whether TBT applies to digital goods without any physical product manifestation. E.g., Oddenino argues that TBT could apply to cybersecurity standards, citing discussions at the TBT Committee. Fleuter argues that digital products as services under WTO rules, which would preclude use of TBT. Oddenino, Alberto. “Digital standardization, cybersecurity issues and international trade law.” *Questions of International Law* 51 (2018): 31; Fleuter, Sam. “The Role of Digital Products Under the WTO: A New Framework for GATT and GATS Classification.” *Chi. J. Int'l L.* 17 (2016): 153.

<sup>68</sup> TBT Agreement, Article 2.10.

<sup>69</sup> These include Brazil, China, India, Singapore, and South Korea. Bütke and Mattli, *The new global rulers*. See too [ISO's webpage featuring national examples of standards used for public policy](#); Winfield, Alan FT, and Marina Jirotko. “Ethical governance is essential to building trust in robotics and artificial intelligence systems.” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, no. 2133 (2018): 20180085.

<sup>70</sup> EU Regulation 1025/2012; See too Mattli, “The politics and economics of international institutional standards setting.”



procurement where they exist.<sup>71</sup> The US Department of Defence (DoD), for instance, uses multiple international product and process standards in its software procurement,<sup>72</sup> and this appears set to continue based on the 2018 U.S. Department of Defense Artificial Intelligence Strategy.<sup>73</sup> Beyond regulatory obligations and procurement requirements, governments spread standards through adoption in their own operations.<sup>74</sup>

National action may threaten a fractured global governance landscape and fears of a race to the bottom in regulatory stringency, including that of standards. In a race to the bottom in regulatory stringency, AI development organizations may, in the future, choose to locate in jurisdictions that impose a lower regulatory burden; these organizations need not actually relocate, or threaten to do so, in order to impose downward pressure on regulatory oversight.<sup>75</sup> National strategies already witnessed have proposed policy changes to encourage AI development. Such national actions will undoubtedly continue and court leading AI development organizations.

WTO institutions, if actively used for the purpose, may be able to moderate these concerns of a race to the bottom. Notably, moreover, the global and concentrated nature of markets for AI and related industries will see MNCs use standards internationally. Analogous to government procurement, MNCs may themselves demand contractors adhere to international standards.<sup>76</sup> Such standards include network-product and network-process standards to ensure an interoperable supply chain.<sup>77</sup> They also include enforced-product and enforced-process

---

<sup>71</sup> “Agreement on Government Procurement”, entered into force January 1, 1995. *United Nations Treaty Series*, v. 1868. <https://treaties.un.org/doc/Publication/UNTS/Volume%201868/v1868.pdf>.

<sup>72</sup> Within DoD, individual managers have discretion if and how to use these standards for particular projects: if a standard is to be used, it is cited in Requests for Proposal, included in subsequent contracts, and then used to evaluate contract compliance. In the case of ISO/IEC/IEEE 15288: Systems and software engineering—System life cycle processes, DoD project managers are directed to tailor requirements to their particular project characteristics by, for example, removing out of scope criteria and assessments. The standard establishes a general framework for the life cycle of an engineered system and then defines a set of processes within the life cycle, and can be tailored to apply to use with hardware, software, data, humans, processes, procedures, facilities, materials, and more. “Acquisition Program Resources.” *Office of the Deputy Assistant Secretary of Defense: Systems Engineering*. March 30, 2017. <https://www.acq.osd.mil/se/apr/apr-4.html>; *Best Practices for Using Systems Engineering Standards (ISO/IEC/IEEE 15288, IEEE 15288.1, and IEEE 15288.2) on Contracts for Department of Defense Acquisition Programs*. Washington, D.C.: Office of the Deputy Assistant Secretary of Defense for Systems Engineering, 2017. <https://www.acq.osd.mil/se/docs/15288-Guide-2017.pdf>.

<sup>73</sup> The plan identifies a goal to “establish key AI building blocks and standards” (p.6) and remains focused on processes: contractors’ prototype solutions should employ “standardized processes with respect to areas such as data, testing and evaluation, and cybersecurity.” (p.9) DoD. “Summary of the Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity.” *DoD*, 2019. <https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF>.

<sup>74</sup> Military bases and municipalities in the U.S. and Europe have adopted the ISO 14001 standard, for example. Prakash, Aseem, and Matthew Potoski. *The voluntary environmentalists*.

<sup>75</sup> Empirically there is more inertia in policy and corporate investment than is often assumed. Races to the bottom in regulatory stringency are not often observed in practice for this reason. Radaelli, Claudio M. “The puzzle of regulatory competition.” *Journal of Public Policy* 24, no. 1 (2004): 1-23. But given the stakes in AI development, such inertia is less likely. Absent international coordination, whether by standards or another method, a race to the bottom may play out over time as countries improve training of AI researchers and fail to provide regulatory oversight.

<sup>76</sup> May, Christopher. “Who’s in Charge? Corporations as Institutions of Global Governance.” *Palgrave Communications* 1, no. 1 (2015): 1.

<sup>77</sup> *Ibid.*; Guler, Isin, Mauro F. Guillén, and John Muir Macpherson. “Global competition, institutions, and the diffusion of organizational practices: The international spread of ISO 9000 quality certificates.” *Administrative science quarterly* 47, no. 2 (2002): 207-232.

standards to meet customer demand.<sup>78</sup> In addition to reduced costs from supply chain interoperability and increased revenues from meeting customer demand, MNCs--and other firms alike--have further incentives to adopt international standards: standards can provide protection from liability in lawsuits and can lower insurance premiums.<sup>79</sup>

Together these mechanisms can be used to encourage movement toward a unified global governance landscape for AI standards. National governments and MNCs can mandate use of standards, product and process, network and enforced alike. WTO rules require consistent use of international product standards globally. The incentives of MNCs encourage consistent use of international standards--both product and process--globally. If a large national market mandates adherence to a standard, MNCs may keep administration costs low by complying across the globe. If they do, then MNCs are incentivized to lobby other jurisdictions to pass similar laws, lest local competition be at an advantage.<sup>80</sup> That means, given that many leading AI research efforts are within MNCs, insofar as one country incorporates international AI standards into local law, others will face pressure to follow suit. This was witnessed, for example, with environmental regulation passed in the U.S., which subsequently led DuPont to lobby for a global agreement to ban ozone-depleting chemicals in order to see its international competition similarly regulated.<sup>81</sup> This phenomenon is currently witnessing the globalization of data protection regulations at the behest of the GDPR. The analysis of global governance mechanisms in this section should not be portrayed as arguing that using these tools to spread and enforce AI standards globally will be easy. But the tools do exist, and concerted efforts to make use of them are a worthy endeavour.

In sum, the scope for standards in the global governance of AI research and development is not predetermined. Recalling our standards typology, the object of standardization and incentives therein will determine particular needs for standards development and complementary institutions. Standards for AI product specification and development processes have numerous precedents, while standards to govern fundamental research approaches are without precedent. More generally, if international experts engage in the standardization process, this serves to legitimize the resulting standard. If states and MNCs undertake efforts to adopt and spread the standard, it will similarly grow in influence. Active institutional entrepreneurship can influence the development of and scope for international standards in AI.

---

<sup>78</sup> May, "Who's in Charge?"; Prakash, Aseem, and Matthew Potoski. "Racing to the bottom? Trade, environmental governance, and ISO 14001." *American journal of political science* 50, no. 2 (2006): 350-364; Vogel, David. *The Market for Virtue: The Potential and Limits of Corporate Social Responsibility*. Washington, D.C.: Brookings Institution Press, 2005.

<sup>79</sup> Cihon, P., Michel-Guitierrez, G., Kee, S., Kleinaltenkamp, M., and T. Voigt. "Why Certify? Increasing adoption of the proposed EU Cybersecurity Certification Framework." Masters thesis, University of Cambridge, 2018; for a proposal to incentivize AI certification via reduced liability in the US context, see too, Scherer, "Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies."

<sup>80</sup> Murphy, Dale D. *The structure of regulatory competition: Corporations and public policies in a global economy*. Oxford: Oxford University Press, 2004; Bradford, Anu. "The Brussels effect." *Nw. UL Rev.* 107 (2012): 1-68.

<sup>81</sup> Murphy, *The structure of regulatory competition*; Hale, Thomas, David Held, and Kevin Young. *Gridlock: Why Global Cooperation Is failing When We Need It Most*. Cambridge: Polity, 2013.

## 3. Current Landscape for AI Standards

### 3A. International developments

Given the mechanisms outlined in the previous section, international standards bodies are a promising forum of engagement for AI researchers. To date, there are two such bodies working on AI: ISO/IEC JTC 1 Standards Committee on Artificial Intelligence (SC 42) and the working groups of IEEE SA's AI standards series. Figure 2 categorizes the standards under development within the externality-object typology, as of January 2019.

**Figure 2. International AI standards under development**

<p style="text-align: center;"><b>Network - Product</b></p> <ul style="list-style-type: none"> <li>● Foundational Standards: Concepts and terminology (SC 42 WD 22989), Framework for Artificial Intelligence Systems Using Machine Learning (SC 42 WD 23053)</li> <li>● Transparency of Autonomous Systems (defining levels of transparency for measurement) (IEEE P7001)</li> <li>● Personalized AI agent specification (IEEE P7006)</li> <li>● Ontologies at different levels of abstraction for ethical design (IEEE P7007)</li> <li>● Wellbeing metrics for ethical AI (IEEE P7010)</li> <li>● Machine Readable Personal Privacy Terms (IEEE P7012)</li> <li>● Benchmarking Accuracy of Facial Recognition systems (IEEE P7013)</li> </ul>	<p style="text-align: center;"><b>Network - Process</b></p> <ul style="list-style-type: none"> <li>● Model Process for Addressing Ethical Concerns During System Design (IEEE P7000)</li> <li>● Data Privacy Process (IEEE P7002)</li> <li>● Methodologies to address algorithmic bias in the development of AI systems (IEEE P7003).</li> <li>● Process of Identifying and Rating the Trustworthiness of News Sources (IEEE P7011)</li> </ul>
<p style="text-align: center;"><b>Enforced - Product</b></p> <ul style="list-style-type: none"> <li>● Certification for products and services in transparency, accountability, and algorithmic bias in systems (IEEE ECPAIS)</li> <li>● Fail-safe design for AI systems (IEEE P7009)</li> </ul>	<p style="text-align: center;"><b>Enforced - Process</b></p> <ul style="list-style-type: none"> <li>● Certification framework for child/student data governance (IEEE P7004)</li> <li>● Certification framework for employer data governance procedures based on GDPR (IEEE P7005)</li> <li>● Ethically Driven AI Nudging methodologies (IEEE P7008)</li> </ul>

SC 42 is likely the more impactful venue for long-term engagement. This is primarily because IEEE standards have fewer levers for adoption than their ISO equivalents. WTO TBT rules can apply to both IEEE and ISO/IEC product standards, but their application to IEEE was only asserted in 2017 and has never been tested.

<sup>82</sup> As discussed in the previous section, ISO standards are mandated in government regulation; similar research could find no such mandates for IEEE standards. Procurement requirements are common for both IEEE and ISO/IEC standards, and market mechanisms similarly encourage both. ISO has had global success with many enforced standards, whereas IEEE has no equivalent experience to date.<sup>83</sup> States have greater influence in ISO/IEC standards development than that of IEEE, and state involvement has enhanced the effectiveness of past standards with enforcement mechanisms.<sup>84</sup> Thus, given that enforcement of ISO/IEC standards has more mechanisms for global reach, participation in ISO/IEC JTC 1 may be more impactful than in IEEE.

Ongoing SC 42 efforts are, so far, few in number and preliminary in nature. (See Appendix 1 for a full list of SC 42 activities.) The most pertinent standards working group within SC 42 today is on Trustworthiness. The Trustworthiness working group is currently drafting three technical reports on robustness of neural networks, bias in AI systems, and an overview of trustworthiness in AI.

IEEE's AI standards are further along than those of SC 42. (See Appendix 2 for a full list of IEEE SA P7000 series activities, as of January 2019.) Work on the series began in 2016 as part of the IEEE's larger Global Initiative on Ethics of Autonomous and Intelligent Systems. IEEE's AI standards series is broad in scope, and continues to broaden with recent additions including a project addressing algorithmic rating of fake news. Of note to AI researchers interested in long-term development should be P7009 Fail-Safe Design of Autonomous and Semi-Autonomous Systems. The standard, under development as of January 2019, includes "clear procedures for measuring, testing, and certifying a system's ability to fail safely."<sup>85</sup> Such a standard, depending on its final scope, could influence both research and development of AI across many areas of focus. Also of note is P7001 Transparency of Autonomous Systems, which seeks to define measures of transparency. Standardized methods and measurements of system transparency could inform monitoring measures in future agreements on advanced AI development.

IEEE SA recently launched the development of an Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS). Unlike the other IEEE AI standards, development is open to paid member organizations, not interested individuals. ECPAIS seeks to develop three separate processes for certifications related to transparency, accountability, and algorithmic bias. ECPAIS is in an early stage, and it remains to be seen to what extent the certifications will be externally verified.<sup>86</sup> Absent an enforcement mechanism, such certifications could be subject to the failings of negative externality standards that lack enforcement mechanisms.<sup>87</sup>

---

<sup>82</sup> "IEEE Position Statement: IEEE Adherence to the World Trade Organization Principles for International Standardization."

<sup>83</sup> IEEE acknowledges that their AI standards are "unique" among their past standards: "Whereas more traditional standards have a focus on technology interoperability, safety and trade facilitation, the IEEE P7000 series addresses specific issues at the intersection of technological and ethical considerations." [IEEE announcement webpage](#).

<sup>84</sup> Vogel, David. "The Private Regulation of Global Corporate Conduct." in Mattli, Walter., and Ngaire. Woods, eds. *The Politics of Global Regulation*. Princeton: Princeton University Press, 2009.

<sup>85</sup> "P7009 Project Authorization Request." *IEEE-SA*. Published July 15, 2017. <https://development.standards.ieee.org/get-file/P7009.pdf?t=93536600003>.

<sup>86</sup> [IEEE announcement webpage](#).

<sup>87</sup> Calo, Ryan. Twitter Thread. October 23, 2018, 1:39PM. <https://twitter.com/johnchavens/status/1054848219618926592>.

### 3B. National priorities

There are three important developments to note for national policies on standards for AI. First, key national actors, including the U.S. and China, agree that international standards in AI are a priority. Second, national strategies for AI also indicate that countries plan to pursue national standards. Third, given the market structure in AI industry, countries are incentivized to ensure that international standards align as closely to national standards as possible.

First, international standards are a stated priority for key governments. The recently released U.S. Executive Order on Maintaining American Leadership in Artificial Intelligence identified U.S. leadership on international technical standards as a priority and directed the National Institute for Standards and Technology to draft a plan to identify standards bodies for the government to engage.<sup>88</sup> The Chinese government has taken a similar position in an AI Standardization White Paper published by the China Electronics Standardization Institute (CESI) within the Ministry of Industry and Information Technology in 2018.<sup>89</sup> The white paper recommended that “China should strengthen international cooperation and promote the formulation of a set of universal regulatory principles and standards to ensure the safety of artificial intelligence technology.”<sup>90</sup> This recommendation was corroborated by previous CESI policies, e.g., its 2017 Memorandum of Understanding with the IEEE Standards Association to promote international standardization.<sup>91</sup>

Second, national standards remain relevant. Historically, observers have argued that Chinese national standards in fields auxiliary to AI, including cloud computing, industrial software, and big data, differ from international standards in order to support domestic industry.<sup>92</sup> These differences have not been challenged under WTO rules. However, these same observers do note that China is increasingly active in international standards activities. In January 2018, China established a national AI standardization group, which will be active with ISO/IEC JTC 1 SC 42 and coordinate some 23 active AI-related national standardization processes, focused on platform/support capabilities and key technologies like natural language processing, human-computer interaction, biometrics, and computer vision.<sup>93</sup>

Historically, the U.S. has also emphasized the importance of standardization for AI without specifying such efforts occur at the international level. The 2016 U.S. National AI Research and Development Strategic Plan,

---

<sup>88</sup> “Executive Order on Maintaining American Leadership in Artificial Intelligence,” February 11, 2019, <https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence>.

<sup>89</sup> China Electronics Standardization Institute (CESI). “AI Standardization White Paper,” 2018, Translation by Jeffrey Ding. [https://docs.google.com/document/d/1VqzyN2KINmKmY7mGke\\_KR77o1XQriwKGsuj9dO4MTDo/](https://docs.google.com/document/d/1VqzyN2KINmKmY7mGke_KR77o1XQriwKGsuj9dO4MTDo/); see too Ding, Jeffrey, Paul Triolo and Samm Sacks. “Chinese Interests Take a Big Seat at the AI Governance Table.” *New America* (2018).

<https://www.newamerica.org/cybersecurity-initiative/digichina/blog/chinese-interests-take-big-seat-ai-governance-table/>.

<sup>90</sup> CESI, “AI Standardization White Paper,” p. 4.

<sup>91</sup> See [IEEE Beyond Standards blog post](#).

<sup>92</sup> Ding, Jeffrey. “Deciphering China’s AI Dream.” *Future of Humanity Institute, University of Oxford* (2018); Wübbecke, Jost, Mirjam Meissner, Max J. Zenglein, Jaqueline Ives, and Björn Conrad. “Made in China 2025: The making of a high-tech superpower and consequences for industrial countries.” *Mercator Institute for China Studies* 17 (2016).

<sup>93</sup> CESI, “AI Standardization White Paper.”

for instance, identified 10 key areas for standardization: software engineering, performance, metrics, safety, usability, interoperability, security, privacy, traceability, and domain-specific standards.<sup>94</sup>

Other countries are also considering national standards. An overview of AI national and regional strategies describes plans for standards from Australia, the Nordic-Baltic Region (Denmark, Estonia, Finland, the Faroe Islands, Iceland, Latvia, Lithuania, Norway, Sweden, and the Åland Islands), and Singapore.<sup>95</sup> The Chief Scientist of Australia has proposed an AI product and process voluntary certification scheme to support consumer trust.<sup>96</sup> Insofar as these national strategies seek to develop AI national champions and given the network effects inherent in AI industry,<sup>97</sup> AI nationalism is of transnational ambition.

This leads to the third important point: national efforts will likely turn to international standards bodies in order to secure global market share for their national champions. Successful elevation of national standards to the international level benefits national firms that have already built compliant systems. Successful inclusion of corporate patents into international standards can mean lucrative windfalls for both the firm and its home country.<sup>98</sup>

If one state seeks to influence international standards, all others have incentive to do similarly, else their nascent national industries may lose out. Given that both the U.S. and China have declared intent to engage in international standardization, this wide international engagement will likely come to pass. One illustrative case of the consequences of failure to follow competitors in international standardization is offered by the U.S. machine tools industry. This industry once described by Ronald Reagan as a “vital component of the U.S. defense base,” did not seek to influence global standards on related products, and has declined precipitously under international competition. This stands in contrast to the standards engagement and continued strength of the sector in Germany and Italy.<sup>99</sup> Furthermore, the WTO rules outlined in *Section 2.C.iii*, if enforced, require that national regulations cite international standards. This means that failure to secure international standards that reflect preexisting national ones could require changes in national regulation to encourage global competition. Thus, such developments could cost national industry both internationally *and* domestically. This means that countries will likely engage in international standards bodies that govern priority industries like AI.

---

<sup>94</sup> National Science and Technology Council. “The National Artificial Intelligence Research and Development Strategic Plan.” *Executive Office of the President of the United States*. (2016). [https://www.nitrd.gov/PUBS/national\\_ai\\_rd\\_strategic\\_plan.pdf](https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf).

<sup>95</sup> Dutton, Tim. “An Overview of National AI Strategies”. *Medium*. Published June 28, 2018. <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>.

<sup>96</sup> Finkel, A. “What will it take for us to trust AI?” *World Economic Forum* <https://www.weforum.org/agenda/2018/05/alan-finkel-turing-certificate-ai-trust-robot/>.

<sup>97</sup> Scale unlocks further user-generated data and enables hiring talent, which in turn both improve the underlying product, which in turn increases users and attracts further talent in a virtuous circle.

<sup>98</sup> See Krasner, Stephen D. “Global communications and national power: Life on the Pareto frontier.” *World politics* 43, no. 3 (1991): 336-366; Drezner, Daniel W. *All politics is global: Explaining international regulatory regimes*. Princeton University Press, 2008.

<sup>99</sup> See Büthe and Mattli, *The new global rulers*, Chapter 6.

### Case of 5G: International standards with implications for national champions

Although telecom standards and standardization bodies differ from those leading in AI standardization, the ongoing development of 5G standards is an illustrative case to consider regarding states' interests. Previous generations of mobile telephony standards did not see a single, uncontested global standard, with Europe and the US on inoperable 3G standards and the LTE standard facing competition before solidifying its global market dominance in 4G. The global economies of scale resulting from 4G standard consolidation may see a uniform standard adopted globally for 5G from the start.<sup>100</sup> This globally integrated market will offer positive-sum outcomes to cooperation, albeit with some countries winning more than others. These incentives for network-product standards may very well be larger than those present in AI.

These incentives are driving participation in efforts at the focal standardization body, 3GPP, which set LTE for 4G as well as some past generation standards, to set the radio standard.<sup>101</sup> At stake in the standardization process is the economic bounty from patents incorporated into the standard and their resulting effects on national industry competitiveness in the global market. One estimate claims that U.S. firm Qualcomm owns approximately 15 percent of 5G patents, with Chinese companies, led by Huawei, controlling about 10 percent.<sup>102</sup> One example of Huawei's success in 5G standards was the adoption of its supported polar coding method for use in control channel communication between end devices and network devices.<sup>103</sup>

In contrast to a positive-sum game with distributional consequences common in international standards, the use of national standards reverts to a protectionist zero-sum game. In the past, there has been criticism of China's efforts to use national standards towards protectionism with requirements that differ from international standards. In 5G and AI standards, however, China has sought to engage in international standards bodies, thereby mitigating this past concern and responding to past international pressure to reduce trade barriers.<sup>104</sup> The Trump administration opposes China's international standards activities, in keeping with its zero-sum perspective on international trade. For example, the Committee on Foreign Investment in the United States (CFIUS) decision to block the foreign acquisition of Qualcomm came out of concern that it "would leave an opening for China to expand its influence on the 5G standard-setting process."<sup>105</sup> No

---

<sup>100</sup> Brake, Doug. "Economic Competitiveness and National Security Dynamics in the Race for 5G between the United States and China." *Information Technology & Innovation Foundation*. (2018).

[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3142229](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3142229). Others see the possibility of a partially bifurcated system. Triolo, Paul and Kevin Allison. "The Geopolitics of 5G." *Eurasia Group*. (2018). <https://www.eurasiagroup.net/live-post/the-geopolitics-of-5g>.

<sup>101</sup> This is a simplified picture, as the ITU governs spectrum allocation for 5G as well as a standards development roadmap.

<sup>102</sup> LexInnova, "5G Network Technology: Patent Landscape Analysis" (2017) cited in Brake "Economic Competitiveness and National Security Dynamics in the Race for 5G."

<sup>103</sup> Brake "Economic Competitiveness and National Security Dynamics in the Race for 5G."

<sup>104</sup> Greenbaum, Eli. "5G, Standard-Setting, and National Security," *Harvard Law National Security Journal* (July, 2018), <http://harvardnsj.org/2018/07/5g-standard-setting-and-national-security/>.

<sup>105</sup> Mir, Aimen N. "Re: CFIUS Case 18-036: Broadcom Limited (Singapore)/Qualcomm Incorporated." Department of the Treasury. Letter, p. 2.

[https://www.sec.gov/Archives/edgar/data/804328/000110465918015036/a18-7296\\_7ex99d1.htm](https://www.sec.gov/Archives/edgar/data/804328/000110465918015036/a18-7296_7ex99d1.htm).

longer does the U.S. view international participation as a way to reduce trade barriers; rather, it sees international participation as a way to shift influence in China's favor globally.

Despite these politics, global standards will improve market efficiency and lead to better outcomes for all. Some will be better off than others, however. The distributional consequences of 5G standards may be larger than those for AI in the short-term, but this case nonetheless has implications for international efforts towards AI standards. The future may see similarly politicized standardization processes for AI. China's formulated policies for international engagement on AI standards will likely see other countries engage in order to encourage a more balanced result. This engagement means that AI researchers' efforts to influence standards will be supported but also that they likely will be increasingly politicized. Yet, to be clear, AI standards are not currently as visible or politicized as telecom standards, which have already seen four previous iterations of standards and the emergence of large globally integrated markets dependent upon them.

### 3C. Private initiatives

In addition to international and national standards, there are a number of private initiatives that seek to serve a standardizing role for AI. Standards, most commonly network-product standards, can arise through market forces. Notable examples include the QWERTY keyboard, dominance of Microsoft Windows, VHS, Blu-Ray, and many programming languages. Such market-driven product standards can produce suboptimal outcomes where proprietary standards are promoted for private gain or standards may fail to spread at all due to a lack of early adopters.<sup>106</sup>

In AI, software packages and development environments, e.g., TensorFlow, PyTorch and OpenAI Gym, are privately created, are used widely, and perform a standardizing role. Market forces can also encourage, though not develop in their own right, network-process and enforced standards through customer demands on MNCs and MNC pressure on their supply chains, as explained in *Section 2.C.iii*. For example, the CleverHans Adversarial Examples Library,<sup>107</sup> if incorporated into an adversarial example process check that became widely adopted in practice, would be such a standard. Another example is Microsoft's Datasheets for Datasets standard practice to report on data characteristics and potential bias that is used across the company.<sup>108</sup> Researchers will continue to develop software packages and benchmarks. This approach does not necessarily require an additional commitment of time beyond their research work, whereas engaging on traditional standards development does require some time commitment. Some of these packages and benchmarks may spread to the extent that they become industry standards. But these standards will face difficulties in securing global dissemination and enforcement. Yet, as discussed in *Section 4* below private standards can be turned into international standards with a concerted effort.

---

<sup>106</sup> Mattli, Walter. "Public and Private Governance in Setting International Standards." In Kahler, Miles, and David A. Lake, eds. *Governance in a global economy: Political authority in transition*. Princeton University Press, 2003.

<sup>107</sup> Papernot, Nicolas, Fartash Faghri, Nicholas Carlini, Ian Goodfellow, Reuben Feinman, Alexey Kurakin, Cihang Xie et al. "Technical report on the cleverhans v2. 1.0 adversarial examples library." *arXiv preprint arXiv:1610.00768* (2016).

<sup>108</sup> Gebu, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Dauméé III, and Kate Crawford. "Datasheets for Datasets." *arXiv:1803.09010* (2018).



Still other groups are developing standards in the broad sense of a guide for judging behavior. The 2017 Asilomar Conference on Beneficial AI yielded a set of AI Principles that address areas of research, ethics and values, and long-term issues, which have been signed by some 1300 AI and robotics researchers as well as 2500 others.<sup>109</sup> Among these principles was a commitment to safety standards in particular: “Teams developing AI systems should actively cooperate to avoid corner-cutting on safety standards.”<sup>110</sup> The Association for Computing Machinery (ACM), a professional association, maintains a Code of Ethics and Professional Conduct for its members. This Code includes many principles, including “Avoid harm.”<sup>111</sup> The ACM has also called for a new peer review standard that requires researchers to acknowledge “negative implications” of their research.<sup>112</sup> The Partnership on AI, a multistakeholder forum founded by leading AI firms, seeks to develop best practices for AI research and development.<sup>113</sup> These standards, broadly defined, do not benefit directly from the dissemination and enforcement mechanisms outlined in *Section 2Ciii*. However, such standards may have normative power in influencing actors who subsequently engage in standardization activities that produce standards which are subject to mechanisms of dissemination and enforcement.

## 4. Recommendations

Today, AI standards development is already underway at both ISO/IEC and IEEE. National strategies, including those of the U.S. and China, prioritize engagement in standardization processes for AI. Thus, the agenda is set. Engagement in these processes today can benefit from these ongoing processes and national foci. As time goes on, however, standards bodies may become increasingly politicized just as multiple iterations of telecom standards have, over time, given rise to highly politicized international tension over 5G. This section offers recommendations to use standards to help support AI policy goals starting today.

### 4A. Engage in ongoing processes

How can the AI community, namely researchers and research organizations, engage effectively? There are four elements necessary for successful influence in international standards bodies:

- technical expertise,
- financial resources,
- timely information, and
- effective institutional knowledge.<sup>114</sup>

---

<sup>109</sup> “Asilomar AI Principles.” [Future of Life Institute website](#).

<sup>110</sup> Ibid.

<sup>111</sup> “ACM Code of Ethics and Professional Conduct.” [Association for Computing Machinery website](#).

<sup>112</sup> Hecht, B., Wilcox, L., Bigham, J.P., Schöning, J., Hoque, E., Ernst, J., Bisk, Y., De Russis, L., Yarosh, L., Anjum, B., Contractor, D. and Wu, C. 2018. It’s Time to Do Something: Mitigating the Negative Impacts of Computing Through a Change to the Peer Review Process. ACM Future of Computing Blog. <https://acm-fca.org/2018/03/29/negativeimpacts/>.

<sup>113</sup> “About Us.” [Partnership on AI website](#).

<sup>114</sup> Büthe and Mattli, *The new global rulers*. I adapt the final element from their focus on “institutional complementarity” that sees domestic institutions that effectively mirror specifically ISO and IEC. My broader reading, effective institutional awareness, enables actors to create new structures or otherwise navigate these challenges of institutional complementarity.

The AI research community already has technical expertise and financial resources, but not up-to-date information on proceedings within standards bodies nor the institutional knowledge required to successfully intervene. The following four recommendations help fill in these gaps.

#### 4Ai. Build capacity for effective engagement

AI researchers are unlikely to have experience engaging national and international standards bodies. Of leading AI organizations, only Google and Microsoft participate in the U.S. standards committee that is affiliated with ISO/IEC JTC 1 SC 42; none participates in the U.K. equivalent. Similarly, IEEE P7000 series working groups see very few volunteers from leading organizations.<sup>115</sup>

In order to successfully influence standardization outcomes, researchers should develop expertise in these standardization processes. In some cases, researchers need not go far to find this expertise. Large firms may already have teams working on creating and complying with international standards, though they may focus more on products as opposed to AI research and development.

Research institutions and firms can learn more about ongoing standardization processes by participating in the Open Community for Ethics in Autonomous and Intelligent Systems (OCEANIS).<sup>116</sup> OCEANIS is a coordinating forum for standards organizations and other interested organizations to discuss efforts to use standards to further development of autonomous and intelligent systems. It was co-founded by the IEEE and IEC, among other national and regional standards bodies. OCEANIS does not produce standards itself, but could be a useful venue for organizations seeking to build capacity prior to engaging directly in standardization processes.

Beyond expertise, perspective matters: it is important to view standards as a policy tool for encouraging positive AI outcomes. Technologies are not apolitical<sup>117</sup> and neither are the processes that shape them.<sup>118</sup> With this understanding, standards are not simply a response to a particular market need, but, more broadly, a tool of global governance. Strategic engagement in standardization now can help direct wider consideration to important areas like AI safety. ISO and IEEE have formalized standards maintenance procedures so that standards can be updated as the state of the art progresses.<sup>119</sup> The important step today is to understand and start using this tool for global governance.

---

<sup>115</sup> This was indicated by observations of one working group and an interview with the chair of another working group.

<sup>116</sup> “Participation.” [OCEANIS website](#).

<sup>117</sup> See, e.g., Lessig, Lawrence. *Code: And Other Laws of Cyberspace, Version 2.0*. New York: Basic Books, 2006.

<sup>118</sup> Büthe and Mattli, *The new global rulers*.

<sup>119</sup> For details on ISO’s systematic review process, see “Guidance on the Systematic Review Process in ISO.” ISO. Published May 2017. [https://www.iso.org/files/live/sites/isoorg/files/store/en/Guidance\\_systematic\\_review.pdf](https://www.iso.org/files/live/sites/isoorg/files/store/en/Guidance_systematic_review.pdf); For details on IEEE’s maintenance process, see “Next Steps Kit: Guidelines for Publication, Recognition Awards and Maintenance.” IEEE Standards Association. N.D.

[https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/next\\_steps\\_kit.pdf](https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/next_steps_kit.pdf).

#### 4Aii. Engage directly in ongoing processes

There are two, related paths to engage with ISO/IEC JTC 1 SC 42. First, researchers should consider joining the group that mirrors SC 42 within their respective national standards body. It is through these national bodies that researchers can influence and directly engage in SC 42. In the United States, this group is InterNational Committee for Information Technology Standards (INCITS) - Artificial Intelligence. Committee membership is open to any U.S. national who is materially affected by related standardization activities.<sup>120, 121</sup> The equivalent committee for the UK is British Standards Institution ART/1.<sup>122</sup>

The second method of engaging SC 42 is to seek appointment to its expert working groups that drafts standards directly. Such appointments are made by national member organizations, so the first engagement strategy will further the second.

The work of SC 42 is in its early stages. Working Group 3 on Trustworthiness, and specifically its ongoing work on a technical report on robustness in neural networks, is likely the highest value area of engagement at this time. At this preliminary stage, however, participation in a national standards body or SC 42 working group can serve to build career capital and institutional knowledge that will be useful in creating further working groups in the future. These efforts could focus on standards related to AI policy goals; some of these possible standards will be discussed below.

IEEE SA P7000 series working groups are open for interested individuals to join. Indeed, the process of joining is much simpler than that of ISO/IEC JTC 1-related work. One simply needs to contact a working group and express interest. In order to participate in developing the Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS), individuals must be affiliated with organizations possessing an IEEE SA Advanced Corporate Membership.

The work of standards within the IEEE SA P7000 series is at varied stages of completeness. Standards earlier in the sequence have approximately a year left in development and standards later in the sequence have more time. This means that interested researchers should consider engaging soon if they are to have an impact in ongoing working groups. Two standards in particular could support AI policy goals outlined above: P7001 Transparency of Autonomous Systems, which seeks to define measures of transparency and P7009 Fail-Safe Design of Autonomous and Semi-Autonomous Systems. Engagement on these standards could help ensure that their respective scopes support the governance of long-term risks.

---

<sup>120</sup> InterNational Committee for Information Technology Standards (INCITS). “New INCITS Technical Committee on Artificial Intelligence - Notice of January 30-31, 2018 Organizational Meeting and Call for Members.” Email, 2018. [https://standards.incits.org/apps/group\\_public/download.php/94314/eb-2017-00698-Meeting-Notice-New-INCITS-T-C-on-Artificial-Intelligence-January30-31-2018.pdf](https://standards.incits.org/apps/group_public/download.php/94314/eb-2017-00698-Meeting-Notice-New-INCITS-T-C-on-Artificial-Intelligence-January30-31-2018.pdf).

<sup>121</sup> Membership fees vary by organizational affiliation, from several hundred to several thousand dollars per year.

<sup>122</sup> See [BSI committee information webpage](#).

#### 4Aiii. Multinational organizations should become liaisons

Another method of engagement is available to *multinational* industry or other membership associations like the Partnership on AI. These groups are eligible for liaison status with ISO/IEC JTC 1 SC 42 at both the Standards Committee level and Working Group level. Although liaisons cannot vote on final standards, they have significant influence. Participation at the Standards Committee level would allow such an organization to propose new standards, comment on draft standards, and nominate experts for Working Groups.<sup>123</sup>

#### 4B. Pursue parallel standards development

An alternative to direct engagement is the parallel development of standards. This could take many forms in practice. Individual organizations, existing working groups at the Partnership on AI, or other ad hoc consortia could develop, i.e., software libraries, measurement benchmarks, or best practices procedures. Once developed, these approaches could then be transferred into international standards to achieve global dissemination and enforcement.

Indeed, there are numerous examples of organizations and even individual firms transferring existing standards into international ISO standards. The C Programming language was developed at Bell Laboratories before being adopted as an international standard by ISO.<sup>124</sup> More recently, Microsoft transferred its Open XML format to ISO,<sup>125</sup> as did Adobe with PDF.<sup>126</sup> Microsoft's effort is illustrative of the potential for one motivated MNC to influence standardization processes: it placed its experts on several national committees that then influenced discussions at the ISO committee.<sup>127</sup> Smaller firms can also have success: Microsoft's Open XML effort followed another ISO-approved open standard that was submitted by a consortium of smaller companies.<sup>128</sup>

IEEE has also created standards and then seen them adopted by ISO/IEC JTC 1 in the past. Similar efforts could be made in the case of specific AI standards, whether from IEEE's P7000 series or from another organization. Indeed, if an organization like the Partnership on AI were to create AI standards, it could apply for status as a Publicly Available Specifications (PAS) Submitter with the ISO/IEC JTC 1. With this status, a standards organization can submit specifications for a vote among national bodies; over one hundred standards have been approved in this process.<sup>129</sup>

---

<sup>123</sup> "ISO/IEC Directives Part 1 Consolidated ISO Supplement, 1.17 Liaisons with other organizations." Ninth edition. Geneva: ISO, 2018. [https://www.iso.org/sites/directives/current/consolidated/index.xhtml#\\_idTextAnchor095](https://www.iso.org/sites/directives/current/consolidated/index.xhtml#_idTextAnchor095).

<sup>124</sup> Kelechava, Brad. "The Origin of ANSI C and ISO C." [ANSI blog post](#).

<sup>125</sup> Kosek, Jirka. "From the office document format battlefield." *IT Professional* 10, no. 3 (2008).

<sup>126</sup> Orion, Egan. "PDF 1.7 is approved as ISO 32000." [The Inquirer blog post \(archived\)](#).

<sup>127</sup> Büthe and Mattli, *The new global rulers*, p. 186.

<sup>128</sup> Kosek, "From the office document format battlefield."

<sup>129</sup> "JTC 1 PAS Transposition Process." [ISO/IEC JTC 1 website](#).

## 4C. Research standards and strategy for development

This paper and its nascent AI standards strategy serves as a call to AI researchers to engage in order to help develop standards as a global governance mechanism for AI. Further research from the AI research community is needed to ensure that standards under development today can encourage positive outcomes for advanced AI. This work could benefit from researchers from across the AI research field as well as forecasting experts. The lines of work are two-fold. First, technical standards desiderata should be developed. Second, specific strategies to see these standards codified and spread globally should then be created.

### 4Ci. Research technical standards desiderata

Ultimately, AI researchers should seek to consolidate AI standards desiderata for their particular area of focus. Some of this work may take place at existing working groups hosted by the Partnership on AI, discussions within individual organizations, or through other ad hoc gatherings. This paper offers two prototype standards that would support AI policy goals: an AI safety process standard and an AI systems capability standard.

The field of AI Safety is young, but preliminary conversations about how to incorporate safety procedures into a standard framework that can reduce risks globally would be a welcome application of existing research. The first step in this process is the distillation of current best practices. However tentative and incomplete, these practices are an improvement over a disregard for safety--if expectations are calibrated correctly. There are numerous labs around the world today with advanced AI ambitions and no focus on safety.<sup>130</sup> Prototype standards could spread a focus on safety and current best practices globally.<sup>131</sup> One such approach could be a process standard that requires researchers to complete a checklist procedure before undertaking research, namely record a precise specification, measures taken to ensure robustness, and methods of assurance.<sup>132</sup> This approach could then serve as a model for future standards and regulation as system capabilities increase. A more developed version could see researchers certify to a series of best practices. Such a certification framework could eventually be linked to a monitoring regime for defined high risk projects. This certification approach would likely see a series of related standards, which is a common practice. One standard would be definitional: defining high risk projects or developing a risk typology of multiple categories, as is used in functional safety standards. Another standard would then identify best practices and mitigation strategies to be followed at each risk threshold. Additional standards could specify monitoring and certification regimes. When realized, this example may see labs obtain third-party certification subject to verification, e.g., via real-time monitoring of researchers'

---

<sup>130</sup> Baum, "A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy."

<sup>131</sup> E.g., Avoiding negative side effects, avoiding reward hacking, scalable oversight, safe exploration, and robustness to distributional change. Amodei, Dario, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. "Concrete problems in AI safety." *arXiv* (2016). <https://arxiv.org/pdf/1606.06565.pdf>; papers referenced within "Safety and Beneficence of Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI)." *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*. 2017.

[https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead\\_safety\\_beneficence\\_v2.pdf](https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_safety_beneficence_v2.pdf).

<sup>132</sup> See Ortega, Pedro A., Vishal Maini, and the DeepMind Safety Team. "Building safe artificial intelligence: specification, robustness, and assurance." *Medium*. Published September 27, 2018.

<https://medium.com/@deepmindsafetyresearch/building-safe-artificial-intelligence-52f5f75058f1>.

use of large amounts of computing hardware. Although such enforcement regimes may be novel, international standards bodies do have experience with safety standards for emerging technologies, as described in *Section 2Bi*.

Another standard for consideration would permit consistent assessment of system capabilities globally. This type of standard could inform above safety standards by assessing the relative danger of a system or it could facilitate international agreements on AI development and deployment in a variety of domains. Insofar as it was incorporated into an international standard, these practices could be spread globally and possibly facilitate future international agreements. ISO has supported similar efforts to combat climate change in furtherance of the Paris Climate Agreement: it has a series of greenhouse gas emissions measurement standards for individual organizations and auditors.<sup>133</sup> In contrast to the organic spread of private benchmarks, international standards can support universal adoption at a point in the future where it may be needed.

Performance benchmarks already exist for particular tasks. The AI Index incorporates these benchmarks and others to report on AI performance and other metrics annually.<sup>134</sup> Notable benchmarks include the ImageNet corpus, which has served as an image recognition benchmark for research globally and helped drive the rise of deep learning research.<sup>135</sup> The General Language Understanding Evaluation (GLUE)<sup>136</sup> may be a similarly impactful benchmark in the field of natural language understanding. GLUE integrates nine distinct natural language understanding tasks into one benchmark.

As systems become more capable, the integration of tasks into holistic benchmarks will continue. Further work is needed to contextualize these growing modular benchmarks in a broader capabilities framework.<sup>137</sup> An integrative approach could benefit from ongoing efforts to map types of intelligence.<sup>138</sup> Such an approach could then serve as the basis for forecasting efforts and international agreements that see universal adoption. Of course, such a measurement standard cannot come ahead of fundamental research.

#### 4Cii. Research strategies for standards in global governance

Although AI safety research continues and these procedures outlined above are not foolproof, thought on how to implement safety processes at scale needs parallel development with technical safety research itself. Understanding such efforts as enforced-process standards, institutions for both agreement *and* enforcement are

---

<sup>133</sup> *Climate Action*. ISO Focus, May-June 2018.

[https://www.iso.org/files/live/sites/isoorg/files/news/magazine/ISOfocus%20\(2013-NOW\)/en/2018/ISOfocus\\_128/ISOfocus\\_128\\_en.pdf](https://www.iso.org/files/live/sites/isoorg/files/news/magazine/ISOfocus%20(2013-NOW)/en/2018/ISOfocus_128/ISOfocus_128_en.pdf).

<sup>134</sup> See Shoham, Yoav, Raymond Perrault, Erik Brynjolfsson, Jack Clark, James Manyika, Juan Carlos Niebles, Terah Lyons, John Etchemendy, Barbara Grosz and Zoe Bauer, "The AI Index 2018 Annual Report", AI Index Steering Committee, Human-Centered AI Initiative, Stanford University, Stanford, CA, December 2018. <http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf>.

<sup>135</sup> Gershgor, Dave. "The data that transformed AI research—and possibly the world." *Quartz*. Published July 26, 2017. <https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/>.

<sup>136</sup> Wang, Alex, Amapreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. "GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding." *arXiv:1804.07461* (2018).

<sup>137</sup> See Hernández-Orallo, José. "Unbridled mental power." *Nature Physics* 15, no. 1 (2019): 106.

<sup>138</sup> Bhatnagar, Sankalp, Anna Alexandrova, Shahar Avin, Stephen Cave, Lucy Cheke, Matthew Crosby, Jan Feyereisl et al. "Mapping intelligence: Requirements and possibilities." In *3rd Conference on "Philosophy and Theory of Artificial Intelligence*, pp. 117-135. Springer, Cham, 2017.

needed. Although ISO/IEC JTC 1 SC 42 may one day offer a promising home for such efforts, initial proof-of-concept work may be done more effectively elsewhere. Ongoing work on monitoring, incentives, and cooperation, at institutions like OpenAI, FHI Center for the Governance of AI, and the Cambridge Centre for Study of Existential Risk, may prove useful in this effort.

For each identified important area of standardization ongoing, as well as the new efforts identified above, a roadmap for global governance should be developed. This roadmap can then be used by institutional entrepreneurs, whether they be individual researchers, organizations, firms, or larger groups. Each particular standards roadmap could begin by answering the following questions: Which firms, organizations, or states may adopt the standard first? Which policy mechanisms will be most useful in spreading these standard more broadly? How should the broader AI research community support these efforts?

As international and other standards bodies initiate standardization efforts in more areas of AI, an important question to address will be to what extent each needs attention from the AI research community. This is a calculus of interests and impact. If a topic of standardization is relevant to policy goals for the governance of advanced AI and actors' incentives may overlook standards' development to these ends, engagement will be warranted. In other cases, e.g., standards for autonomous vehicles, actors' incentives are aligned so as to not necessitate engagement.<sup>139</sup> Each roadmap should similarly address this question of differential impact.

More broadly, additional research on the politics of standard-setting and standard enforcement is needed. Existing literature focuses primarily on the politics of standard-setting.<sup>140</sup> This work does not focus on standards for digital technologies, however. Furthermore, little work has been done to understand the role of individual firms in setting international standards.<sup>141</sup> Similarly little research has been done on the ways in which standards spread globally in practice. *Section 2Ciii* compiled a series of institutional mechanisms for dissemination and enforcement that warrant further research to analyze their relative performance as well as the influence of global and domestic politics in their processes. This understanding can then inform strategies to spread standards for AI governance.

#### 4D. Use standards as a tool for culture change

Standards can be used to spread a culture of safety<sup>142</sup> and responsibility in AI research and development. Standards can achieve this in four ways.<sup>143</sup> First, the criteria described within a standard set rules and establish

---

<sup>139</sup> No actor wishes to promulgate standards that endanger their own citizens or customers. Insofar as standards for specific AI applications, such as autonomous vehicles, remain national, some efficiency will be lost. But internationally fractured end-application product standards are unlikely to have consequences for desirable openness in research.

<sup>140</sup> See, e.g., Büthe and Mattli, *The new global rulers*; Hallström, *Organizing International Standardization*; Prakash, Aseem, and Matthew Potoski. *The voluntary environmentalists*.

<sup>141</sup> The author's correspondence with a standards scholar indicates that there are no scholarly works focused on corporate lobbying of ISO/IEC or IEEE.

<sup>142</sup> This has also been called a "safety mindset." See "AI Safety Mindset." [Arbital blog](#).

<sup>143</sup> See Scott, W. Richard. *Institutions and Organizations: Ideas, Interests & Identities*. Fourth ed. Foundations for Organizational Science. Los Angeles, California, 2014; Thornton, Patricia H., William. Ocasio, and Michael. Lounsbury. *The Institutional Logics Perspective: A New Approach to Culture, Structure and Process*. Oxford: Oxford University Press, 2012.



expectations. For example, in adopting a transparency standard, an organization commits to the importance of transparency for AI systems. Second, standards establish and reinforce a relational system that see individual researchers and AI development organizations embedded in a larger network. In adopting an international standard, an organization voluntarily acknowledges that outside actors have a stake in the procedures undertaken within the organization. Third, in order to follow the adopted standards, researchers will necessarily carry out practices, repeatedly performing, and internalizing as routine, a culture of responsibility and safety. Fourth, standards will often be embedded directly within products and software packages; individuals' interactions with these artefacts reinforce a culture of safety and responsibility. For example, consider a safety checklist is embedded into a software package that prompts a researcher to address safety risks and mitigation strategies before she can train a model. Regardless of who uses the system, that interaction will reinforce safety.

Understood in this way, standards can be yet another tool for institutional entrepreneurs who promote a culture of responsibility and safety in AI development. Within companies, closer connections between product teams with experience in standards and AI research teams can spread this culture. The adoption of AI standards under development as well as possible future standards can further serve to support this connection within and among AI labs. Outside of a particular company, standards can drive the adoption of best practices more widely across the industry. They can also be bundled with other advocacy efforts that reward responsible labs with better public opinion and access to talented researchers. A culture change is not easy, but standards can help in this path.

## 5. Conclusion

This paper has sought to reframe international standards as tools of AI policy. Some AI policy challenges, e.g., trust among developers and safe ground rules in international competition, warrant global solutions. International standards bodies produce expertise- and consensus-based policies that can provide these solutions. A series of mechanisms can then spread and enforce these policies across the globe. International standards bodies are currently developing AI standards and states have prioritized engagement. The agenda is set, but further expert engagement is needed. This paper has made the case for this engagement, provided an overview of ongoing standardization efforts, and offered detailed recommendations for those who wish to get involved.



## References

- Abbott, Kenneth W., and Duncan Snidal. "International 'standards' and international governance." *Journal of European Public Policy* 8, no. 3 (2001): 345.
- "Acquisition Program Resources." *Office of the Deputy Assistant Secretary of Defense: Systems Engineering*. March 30, 2017. <https://www.acq.osd.mil/se/apr/apr-4.html>.
- "Agreement on Government Procurement", entered into force January 1, 1995. *United Nations Treaty Series*, v. 1868. <https://treaties.un.org/doc/Publication/UNTS/Volume%201868/v1868.pdf>.
- Amodei, Dario, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. "Concrete problems in AI safety." *arXiv* (2016). <https://arxiv.org/pdf/1606.06565.pdf>.
- Armstrong, S., Bostrom, N. & Shulman, C. "Racing to the precipice: a model of artificial intelligence development", Technical Report #2013-1, Future of Humanity Institute, Oxford University: pp. 1-8. (2013). <https://www.fhi.ox.ac.uk/wp-content/uploads/Racing-to-the-precipice-a-model-of-artificial-intelligence-development.pdf>.
- Awad, Edmond, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. "The moral machine experiment." *Nature* 563, no. 7729 (2018): 59.
- Baum, Seth. "A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy." *Global Catastrophic Risk Institute Working Paper 17-1*. (2017). [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3070741](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3070741).
- Best Practices for Using Systems Engineering Standards (ISO/IEC/IEEE 15288, IEEE 15288.1, and IEEE 15288.2) on Contracts for Department of Defense Acquisition Programs*. Washington, D.C.: Office of the Deputy Assistant Secretary of Defense for Systems Engineering, 2017. <https://www.acq.osd.mil/se/docs/15288-Guide-2017.pdf>.
- Bhatnagar, Sankalp, Anna Alexandrova, Shahar Avin, Stephen Cave, Lucy Cheke, Matthew Crosby, Jan Feyereisl et al. "Mapping intelligence: Requirements and possibilities." In *3rd Conference on "Philosophy and Theory of Artificial Intelligence*, pp. 117-135. Springer, Cham, 2017.
- Bostrom, Nick. *Superintelligence*. Oxford: Oxford University Press, 2014.
- "Strategic implications of openness in AI development." *Global Policy* 8, no. 2 (2017): 146.
- Bostrom, Nick, Dafoe, Allan and Carrick Flynn. "Public Policy and Superintelligent AI: A Vector Field Approach." (working paper, Future of Humanity Institute, 2018), <https://nickbostrom.com/papers/aipolicy.pdf>.
- Bradford, Anu. "The Brussels effect." *Nw. UL Rev.* 107 (2012): 1-68.

Brake, Doug. "Economic Competitiveness and National Security Dynamics in the Race for 5G between the United States and China." *Information Technology & Innovation Foundation*. (2018).  
[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3142229](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3142229).

Brundage, Miles, et al. "The malicious use of artificial intelligence: Forecasting, prevention, and mitigation." *Future of Humanity Institute and the Centre for the Study of Existential Risk*. (2018)  
<https://arxiv.org/pdf/1802.07228.pdf>

Brunsson, Nils and Bengt Jacobsson. "The contemporary expansion of standardization" in *A World of Standards*. Oxford: Oxford University Press, 2000, 1-18.

----- "The pros and cons of standardization—an epilogue" in *A World of Standards*. Oxford: Oxford University Press, 2000, 169-173.

Bughin, Jacques, Jeongmin Seong, James Manyika, et al. "Notes from the AI Frontier: Modeling the Impact of AI on the World Economy." *McKinsey Global Institute*. (2018).  
<https://www.mckinsey.com/~media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.ashx>.

Büthe, Tim, and Walter Mattli. *The new global rulers: The privatization of regulation in the world economy*. Princeton University Press, 2011.

Cave, Stephen, and Seán S. ÓhÉigeartaigh. "Bridging near-and long-term concerns about AI." *Nature Machine Intelligence* 1, no. 1 (2019): 5-6.

China Electronics Standardization Institute (CESI). "AI Standardization White Paper," 2018, Translation by Jeffrey Ding.  
[https://docs.google.com/document/d/1VqzyN2KINmKmY7mGke\\_KR77o1XQriwKGsuj9dO4MTDo/edit](https://docs.google.com/document/d/1VqzyN2KINmKmY7mGke_KR77o1XQriwKGsuj9dO4MTDo/edit)

Cihon, Peter. "Regulatory Dynamics of Artificial Intelligence Global Governance." *Typhoon Consulting*. (2018).  
<http://www.typhoonconsulting.com/wp-content/uploads/2018/07/18.07.11-AI-Global-Governance-Peter-Cihon.pdf>

Cihon, Peter, Glenda Michel-Guitierrez, Sam Kee, Moritz Kleinaltenkamp, and Thanel Voigt. "Why Certify? Increasing adoption of the proposed EU Cybersecurity Certification Framework." Masters thesis, University of Cambridge, 2018.

*Climate Action*. ISO Focus, May-June 2018.  
[https://www.iso.org/files/live/sites/isoorg/files/news/magazine/ISOfocus%20\(2013-NOW\)/en/2018/ISOfocus\\_128/ISOfocus\\_128\\_en.pdf](https://www.iso.org/files/live/sites/isoorg/files/news/magazine/ISOfocus%20(2013-NOW)/en/2018/ISOfocus_128/ISOfocus_128_en.pdf).

Dafoe, Allan. "AI Governance: A Research Agenda." *Future of Humanity Institute, University of Oxford*. (2018).

Ding, Jeffrey. "Deciphering China's AI Dream." *Future of Humanity Institute, University of Oxford*. (2018).

Ding, Jeffrey, Paul Triolo and Samm Sacks. "Chinese Interests Take a Big Seat at the AI Governance Table." *New America*. (2018).  
<https://www.newamerica.org/cybersecurity-initiative/digichina/blog/chinese-interests-take-big-seat-ai-governance-table/>.

DoD. "Summary of the Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity." *DoD*, 2019.  
<https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF>

Drezner, Daniel W. *All politics is global: Explaining international regulatory regimes*. Princeton University Press, 2008.

Dutton, Tim. "An Overview of National AI Strategies". *Medium*. Published June 28, 2018.  
<https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>.

"Executive Order on Maintaining American Leadership in Artificial Intelligence," February 11, 2019,  
<https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence>.

Finkel, A. "What will it take for us to trust AI?" *World Economic Forum*. (2018).  
<https://www.weforum.org/agenda/2018/05/alan-finkel-turing-certificate-ai-trust-robot/>.

Fleuter, Sam. "The Role of Digital Products Under the WTO: A New Framework for GATT and GATS Classification." *Chi. J. Int'l L.* 17 (2016): 153.

Geburu, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumeé III, and Kate Crawford. "Datasheets for Datasets." *arXiv:1803.09010* (2018).

Gershgorn, Dave. "The data that transformed AI research—and possibly the world." *Quartz*. Published July 26, 2017. <https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/>.

Grace, Katja, John Salvatier, Allan Dafoe, Baobao Zhang, and Owain Evans. "When will AI exceed human performance? Evidence from AI experts." *Journal of Artificial Intelligence Research* 62 (2018): 729-754.

Greenbaum, Eli. "5G, Standard-Setting, and National Security," *Harvard Law National Security Journal* (July, 2018), <http://harvardnsj.org/2018/07/5g-standard-setting-and-national-security/>.

"Guidance on the Systematic Review Process in ISO." *ISO*. Published May 2017.  
[https://www.iso.org/files/live/sites/isoorg/files/store/en/Guidance\\_systematic\\_review.pdf](https://www.iso.org/files/live/sites/isoorg/files/store/en/Guidance_systematic_review.pdf).

Guler, Isin, Mauro F. Guillén, and John Muir Macpherson. "Global competition, institutions, and the diffusion of organizational practices: The international spread of ISO 9000 quality certificates." *Administrative science quarterly* 47, no. 2 (2002): 207-232.

Hägel, Peter. "Global Governance." *International Relations*. Oxford: Oxford University Press, 2011.

Hale, Thomas, and David Held. *Beyond Gridlock*. Cambridge: Polity. 2017.

Hale, Thomas, David Held, and Kevin Young. *Gridlock: Why Global Cooperation Is failing When We Need It Most*. Cambridge: Polity, 2013.

Hallström, Kristina Tamm. *Organizing International Standardization : ISO and the IASC in Quest of Authority*. Cheltenham: Edward Elgar, 2004.

Hecht, B., Wilcox, L., Bigham, J.P., Schöning, J., Hoque, E., Ernst, J., Bisk, Y., De Russis, L., Yarosh, L., Anjum, B., Contractor, D. and Wu, C. 2018. It's Time to Do Something: Mitigating the Negative Impacts of Computing Through a Change to the Peer Review Process. ACM Future of Computing Blog. <https://acm-fca.org/2018/03/29/negativeimpacts/>.

Hernández-Orallo, José. "Unbridled mental power." *Nature Physics* 15, no. 1 (2019): 106.

IAEA. *Implementing Digital Instrumentation and Control Systems in the Modernization of Nuclear Power Plants*. Vienna: IAEA. 2009. [https://www-pub.iaea.org/MTCD/Publications/PDF/Pub1383\\_web.pdf](https://www-pub.iaea.org/MTCD/Publications/PDF/Pub1383_web.pdf).

"IEEE Position Statement: IEEE Adherence to the World Trade Organization Principles for International Standardization." *IEEE*. Published May 22, 2017. <http://globalpolicy.ieee.org/wp-content/uploads/2017/05/IEEE16029.pdf>.

InterNational Committee for Information Technology Standards (INCITS). "New INCITS Technical Committee on Artificial Intelligence - Notice of January 30-31, 2018 Organizational Meeting and Call for Members." Email, 2018. [https://standards.incits.org/apps/group\\_public/download.php/94314/eb-2017-00698-Meeting-Notice-New-INCITS-TC-on-Artificial-Intelligence-January30-31-2018.pdf](https://standards.incits.org/apps/group_public/download.php/94314/eb-2017-00698-Meeting-Notice-New-INCITS-TC-on-Artificial-Intelligence-January30-31-2018.pdf).

"ISO/IEC Directives Part 1 Consolidated ISO Supplement." Ninth edition. Geneva: ISO, 2018. <https://www.iso.org/sites/directives/current/consolidated/index.xhtml>.

"The ISO Survey of Management System Standard Certifications - 2017 - Explanatory Note. *ISO*. Published August, 2018. [https://isotc.iso.org/livelink/livelink/fetch/-8853493/8853511/8853520/18808772/00.Overall\\_results\\_and\\_explanatory\\_note\\_on\\_2017\\_Survey\\_results.pdf?nodeid=19208898&vernum=-2](https://isotc.iso.org/livelink/livelink/fetch/-8853493/8853511/8853520/18808772/00.Overall_results_and_explanatory_note_on_2017_Survey_results.pdf?nodeid=19208898&vernum=-2).

Jacobsson, Bengt. "Standardization and expert knowledge" in *A World of Standards*. Oxford: Oxford University Press, 2000, 40-50.

Koppell, Jonathan G. S. *World Rule : Accountability, Legitimacy, and the Design of Global Governance*. Chicago: University of Chicago Press, 2010.

Kosek, Jirka. "From the office document format battlefield." *IT Professional* 10, no. 3 (2008).

Krasner, Stephen D. "Global communications and national power: Life on the Pareto frontier." *World politics* 43, no. 3 (1991): 336-366

- Lessig, Lawrence. *Code: And Other Laws of Cyberspace, Version 2.0*. New York: Basic Books, 2006.
- May, Christopher. "Who's in Charge? Corporations as Institutions of Global Governance." *Palgrave Communications* 1, no. 1 (2015): 1.
- Mattli, Walter. "The politics and economics of international institutional standards setting: an introduction." *Journal of European Public Policy* 8, no. 3 (2001): 328-344.
- . "Public and Private Governance in Setting International Standards." In Kahler, Miles, and David A. Lake, eds. *Governance in a global economy: Political authority in transition*. Princeton University Press, 2003.
- Mir, Aimen N. "Re: CFIUS Case 18-036: Broadcom Limited (Singapore)/Qualcomm Incorporated." Department of the Treasury. Letter. (2018).  
[https://www.sec.gov/Archives/edgar/data/804328/000110465918015036/a18-7296\\_7ex99d1.htm](https://www.sec.gov/Archives/edgar/data/804328/000110465918015036/a18-7296_7ex99d1.htm).
- Murphy, Craig N., and JoAnne Yates. *The International Organization for Standardization (ISO): global governance through voluntary consensus*. London: Routledge, 2009.
- Murphy, Dale D. *The structure of regulatory competition: Corporations and public policies in a global economy*. Oxford: Oxford University Press, 2004.
- National Science and Technology Council. "The National Artificial Intelligence Research and Development Strategic Plan." *Executive Office of the President of the United States*. (2016).  
[https://www.nitrd.gov/PUBS/national\\_ai\\_rd\\_strategic\\_plan.pdf](https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf).
- "Next Steps Kit: Guidelines for Publication, Recognition Awards and Maintenance." *IEEE Standards Association*. N.d.  
[https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/next\\_steps\\_kit.pdf](https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/next_steps_kit.pdf).
- Oddenino, Alberto. "Digital standardization, cybersecurity issues and international trade law." *Questions of International Law* 51 (2018): 31.
- Ortega, Pedro A., Vishal Maini, and the DeepMind Safety Team. "Building safe artificial intelligence: specification, robustness, and assurance." *Medium*. Published September 27, 2018.  
<https://medium.com/@deepmindsafetyresearch/building-safe-artificial-intelligence-52f5f75058f1>.
- "P7009 Project Authorization Request." *IEEE-SA*. Published July 15, 2017.  
<https://development.standards.ieee.org/get-file/P7009.pdf?t=93536600003>.
- Papernot, Nicolas, Fartash Faghri, Nicholas Carlini, Ian Goodfellow, Reuben Feinman, Alexey Kurakin, Cihang Xie et al. "Technical report on the cleverhans v2. 1.0 adversarial examples library." *arXiv preprint arXiv:1610.00768* (2016).
- Prakash, Aseem, and Matthew Potoski. *The voluntary environmentalists: Green clubs, ISO 14001, and voluntary environmental regulations*. Cambridge University Press, 2006.

----- "Racing to the bottom? Trade, environmental governance, and ISO 14001." *American journal of political science* 50, no. 2 (2006): 350-364

Radaelli, Claudio M. "The puzzle of regulatory competition." *Journal of Public Policy* 24, no. 1 (2004): 1-23.

Rajchel, Lisa. *25 years of ISO/IEC JTC 1*. ISO Focus+, 2012.

[https://www.iso.org/files/live/sites/isoorg/files/news/magazine/ISO%20Focus%2b%20\(2010-2013\)/en/2012/ISO%20Focus%2b%2c%20June%202012.pdf](https://www.iso.org/files/live/sites/isoorg/files/news/magazine/ISO%20Focus%2b%20(2010-2013)/en/2012/ISO%20Focus%2b%2c%20June%202012.pdf).

"Safety and Beneficence of Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI)." *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*. (2017).

[https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead\\_safety\\_beneficence\\_v2.pdf](https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_safety_beneficence_v2.pdf).

Scherer, Matthew U. "Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies." *Harv. JL & Tech.* 29 (2015): 353-400.

Scott, W. Richard. *Institutions and Organizations: Ideas, Interests & Identities*. Fourth ed. Foundations for Organizational Science. Los Angeles, California, 2014.

Shoham, Yoav, Raymond Perrault, Erik Brynjolfsson, Jack Clark, James Manyika, Juan Carlos Niebles, Terah Lyons, John Etchemendy, Barbara Grosz and Zoe Bauer, "The AI Index 2018 Annual Report", AI Index Steering Committee, Human-Centered AI Initiative, Stanford University, Stanford, CA, December 2018.

<http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf>.

Smith, David J., and Kenneth GL Simpson. *Safety critical systems handbook: a straight forward guide to functional safety, IEC 61508 (2010 Edition) and related standards, including process IEC 61511 and machinery IEC 62061 and ISO 13849*. Elsevier, 2010.

*Technical Barriers to Trade: Reducing trade friction from standards and regulations*. Geneva: WTO, 2015.

[https://www.wto.org/english/thewto\\_e/20y\\_e/tbt\\_brochure2015\\_e.pdf](https://www.wto.org/english/thewto_e/20y_e/tbt_brochure2015_e.pdf).

Thornton, Patricia H., William. Ocasio, and Michael. Lounsbury. *The Institutional Logics Perspective: A New Approach to Culture, Structure and Process*. Oxford: Oxford University Press, 2012.

Triolo, Paul and Kevin Allison. "The Geopolitics of 5G." *Eurasia Group*. (2018).

<https://www.eurasiagroup.net/live-post/the-geopolitics-of-5g>.

Vogel, David. *The Market for Virtue: The Potential and Limits of Corporate Social Responsibility*. Washington, D.C.: Brookings Institution Press, 2005.

----- "The Private Regulation of Global Corporate Conduct." in Mattli, Walter., and Ngaire. Woods, eds. *The Politics of Global Regulation*. Princeton: Princeton University Press, 2009.

Wang, Alex, Amapreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. "GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding." *arXiv:1804.07461* (2018).

Whittaker, Meredith, Kate Crawford, Roel Dobbe, et al. "AI Now Report 2018." *AI Now*. (2018). [https://ainowinstitute.org/AI\\_Now\\_2018\\_Report.pdf](https://ainowinstitute.org/AI_Now_2018_Report.pdf).

Wijkström, Erik, and Devin McDaniels. "Improving Regulatory Governance: International Standards and the WTO TBT Agreement." *Journal of World Trade* 47, no. 5 (2013): 1013-046.

Winfield, Alan FT, and Marina Jirotko. "Ethical governance is essential to building trust in robotics and artificial intelligence systems." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, no. 2133 (2018): 20180085.

Wübbecke, Jost, Mirjam Meissner, Max J. Zenglein, Jaqueline Ives, and Björn Conrad. "Made in China 2025: The making of a high-tech superpower and consequences for industrial countries." *Mercator Institute for China Studies* 17 (2016).

Zwetsloot, Remco and Allan Dafoe. "Thinking About Risks From AI: Accidents, Misuse and Structure." *Lawfare*. (2019). <https://www.lawfareblog.com/thinking-about-risks-ai-accidents-misuse-and-structure>.

## Appendix 1: ISO/IEC JTC 1 SC 42 Ongoing Work<sup>144</sup>

- Working Group 1: Foundational Standards. WG1 has two standards working drafts:
  - WD 22989: [Artificial intelligence -- Concepts and terminology](#)
  - WD 23053: [Framework for Artificial Intelligence \(AI\) Systems Using Machine Learning \(ML\)](#)
  - Both of the above appear to be initial definitional standards. No additional documentation is publically available.
- Working Group 2: Big Data. WG2 incorporates previously ongoing efforts that were assigned to SC 42 at its inception.
  - [ISO/IEC 20546: Information technology — Big data — Overview and vocabulary](#)
  - [ISO/IEC TR 20547-1: Information technology — Big data reference architecture — Part 1: Framework and application process](#)
  - [ISO/IEC TR 20547-1: Information technology — Big data reference architecture — Part 2: Use cases and derived requirements](#) (Published)
  - [ISO/IEC DIS 20547-3: Information technology — Big data reference architecture — Part 3: Reference architecture](#)
  - ISO/IEC DIS 20547-4 Part 4 is managed by JTC 1 SC 27 IT Security techniques
  - [ISO/IEC DIS 20547-3: Information technology — Big data reference architecture — Part 5: Reference architecture](#) (Published)
- Working Group 3: Trustworthiness. WG3 is not currently drafting standards, but is pursuing three technical reports (TR):
  - TR on Bias in AI systems and AI aided decision making
  - TR on Overview of trustworthiness in Artificial Intelligence
  - TR on Assessment of the robustness of neural networks – Part 1: Overview
- Working Group 4: Use Cases and Applications. WG4 is not currently drafting standards, but is pursuing one TR:
  - TR on Artificial Intelligence: use cases

---

<sup>144</sup> For up-to-date information, see the SC 42 blog, [here](#).



## Appendix 2: IEEE AI Standards Ongoing Work

- P7000: [Model Process for Addressing Ethical Concerns During System Design](#)
  - Creates a process model for ethics considerations across development stages.
- P7001: [Transparency of Autonomous Systems](#)
  - Defines levels of measurement for transparency for use during system development.
- P7002: [Data Privacy Process](#)
  - Establishes privacy process management standard to enable conformity assessments.
- P7003: [Algorithmic Bias Consideration](#)
  - Creates a certification framework of methodologies to address negative bias in algorithms.
- P7004: [Child and Student Data Governance](#)
  - Defines a certification framework of methodologies for access, collection, use, storage, sharing, and destruction of child and student data.
- P7005: [Employer Data Governance](#)
  - Establishes a certification framework of methodologies for access, collection, use, storage, sharing, and destruction of employee data.
- P7006: [Personal Data AI Agent Working Group](#)
  - “[D]escribes the technical elements required to create and grant access to a personalized Artificial Intelligence (AI) that will comprise inputs, learning, ethics, rules and values controlled by individuals.” The Project Authorization Request states that a “key goal” of the standard “is to educate government and commercial actors” on the advantages of personalized AI agents.
- P7007: [Ontological Standard for Ethically driven Robotics and Automation Systems](#)
  - Establishes ontologies for ethical design considerations at different levels of abstraction.
- P7008: [Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems](#)
  - Defines common behavior nudges and ethical methodologies for their design.
- P7009: [Fail-Safe Design of Autonomous and Semi-Autonomous Systems](#)
  - Creates technical baseline of methodologies for the design of fail-safe mechanisms.
- P7010: [Wellbeing Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems](#)
  - Establishes metrics for measuring human well-being impacted by systems as well as a related baseline for measurement data.
- P7011: [Process of Identifying & Rating the Trustworthiness of News Sources](#)
  - Provides semi-autonomous processes for rating factual accuracy of news.
- P7012: [Machine Readable Personal Privacy Terms](#)
  - Provides means for individuals to proffer their privacy terms so that they can be machine read by other entities.
- P7013: [Benchmarking Accuracy, Increasing Transparency, and Governing Use of Automated Facial Analysis Technology](#)
  - Establishes demographic definitions and reporting protocols for assessing system performance.
- [The Ethics Certification Program for Autonomous and Intelligent Systems \(ECPAIS\)](#)
  - Certification methodologies for transparency, accountability, and algorithmic bias. Open to IEEE SA Advanced Corporate Members only.