SHOULD WE TRUST A BLACK BOX TO SAFEGUARD HUMAN RIGHTS? A COMPARATIVE ANALYSIS OF AI GOVERNANCE

Scott J. Shackelford*, Isak Nti Asare**, Rachel Dockery***, Angie Raymond****, & Alexandra Sergueeva*****

Abstract

The race to take advantage of the numerous economic, security, and social opportunities made possible by artificial intelligence (AI) is on with nations, intergovernmental organizations, cities, and firms publishing an array of AI strategies. Simultaneously, there are various efforts to identify and distill an array of AI norms. Thus far, there has been limited effort to mine existing AI strategies to see whether common AI norms such as transparency, human-centered design, accountability, awareness, and public benefit are entering into these strategies. Such data is vital to identify areas of convergence and divergence that could highlight opportunities for further norm development in this space by crystallizing State practice.

This Article analyzes more than forty existing national AI strategies paying particular attention to the U.S. context, and then comparing those strategies with private-sector efforts and addressing common criticisms of this process within a polycentric framework. Our findings support the contention that State practices are converging around certain AI principles, focusing primarily upon public benefit. AI is a critical component of international peace, security, and sustainable development in the twenty-first century, and as such

reaching consensus on AI governance will become vital to help build bridges and trust.

Table of Contents

INTRODUCTION

The race to take full advantage of the economic, social, strategic, and political opportunities made possible by artificial intelligence (AI) is picking up pace.[1] From using AI to diagnose COVID-19 by listening to how people are talking[2] and tracking the spread of the pandemic,[3] to designing the next generation of "smart" weapons,[4] jurisdictions from small towns to nations are strategizing how to best make use of AI. The stakes are high with varying approaches to harness this technology being attempted around the world.[5] The winner(s) will enjoy not just a massive first mover advantage, but potentially AI dominance for years, or even

* Executive Director, Ostrom Workshop; Chair, IU-Bloomington Cybersecurity Program; Associate Professor of Business Law and Ethics, Indiana University Kelley School of Business. Special thanks are owed to Noah Halloway and Jalyn Rhodes for their invaluable research support on this Article.
** Associate Director, Language Workshop; Associate Director, Cybersecurity and Global Policy Program.
*** Executive Director, IU Cybersecurity Clinic; Research Fellow in Cybersecurity Law, Indiana University Maurer School of Law.
**** Director, Ostrom Workshop Program on Data and Information Governance; Associate Professor of Business Law and Ethics, Indiana University Kelley School of Business.
***** M.S. in Cybersecurity Risk Management, Indiana University.
[1] *See, e.g.*, Bernard Marr, *What Is The Difference Between Artificial Intelligence And Machine Learning?*, FORBES (Dec. 6, 2016), https://www.forbes.com/sites/bernardmarr/2016/12/06/what-is-the-difference-between-artificial-intelligence-and-machine-learning/#69101912742b (noting that "Artificial Intelligence is the broader concept of machines being able to carry out tasks in a way that we would consider "smart" [while] Machine Learning is a current application of AI based around the idea that we should really just be able to give machines access to data and let them learn for themselves.").
[2] Aaron Holmes, *Do I Sound Sick to You? Researchers are Building AI that Would Diagnose COVID-19 by Listening to People Talk*, BUS. INSIDER (Apr. 30, 2020), https://www.businessinsider.com/ai-labs-diagnose-covid-19-voice-listening-talk-2020-4.
[3] Thomas Macaulay, *AI Model Predicts the Coronavirus Pandemic will End in December*, NEXT WEB (Apr. 29, 2020), https://thenextweb.com/neural/2020/04/29/ai-model-predicts-the-coronavirus-pandemic-will-end-in-december/.
[4] Gordon Cooke, *Magic Bullets: The Future of Artificial Intelligence in Weapons Systems*, U.S. ARMY (June 11, 2019), https://www.army.mil/article/223026/magic_bullets_the_future_of_artificial_intelligence_in_weapons_systems; Kris Osborn, *The U.S. Army's Next Generation of Super Weapons Are Coming*, NAT'L INTEREST (Sept. 16, 2019), https://nationalinterest.org/blog/buzz/us-armys-next-generation-super-weapons-are-coming-80886.
[5] *See* Adrian Pecotic, *Whoever Predicts the Future Will Win the AI Arms Race*, FOREIGN POL'Y (Mar. 5, 2019), https://foreignpolicy.com/2019/03/05/whoever-predicts-the-future-correctly-will-win-the-ai-arms-race-russia-china-united-states-artificial-intelligence-defense/.

decades to come. As Russia's Vladimir Putin proclaimed: "Whoever becomes the leader in this sphere will become the ruler of the world."[6] Although such sentiments may turn out to be alarmist hyperbole,[7] the race to take advantage of the numerous economic, security, and social opportunities made possible by AI is real with nations, intergovernmental organizations, cities, and firms publishing an array of AI strategies. Simultaneously, there are various efforts to identify and distill AI norms to help develop a code of conduct and avoid some of the most destabilizing outcomes.[8] Thus far, there has not yet been an effort to mine existing AI strategies for common AI norms such as transparency, accountability, security, privacy, fairness, human-centered design, and public benefit, and analyze how these are being operationalized through national and local policies. Such data is vital to identify areas of convergence and divergence that could, in turn, highlight opportunities for further norm development in this space by crystallizing State practice. This Article makes an original contribution by analyzing more than forty existing national AI strategies and addressing common criticisms of this process within a polycentric framework. It is the first attempt to analyze AI strategies and norm building through such a comparative lens. AI is a critical component of both international peace and security, and sustainable development, in the twenty-first century, as such reaching consensus on AI governance will become vital to help build bridges, and trust.

The Article is structured as follows. Part 1 unpacks the role of nations in AI governance, tracking parallels with Internet governance and national cybersecurity strategies, with a focus on the U.S. National AI Strategy. Part 2 then moves on to analyze AI strategies from both the public and private sectors paying particular attention to how they treat the topics of transparency, accountability, human-centered design, awareness, and public benefit. Part 3 then explores the implications of our findings for policymakers, assessing criticisms and discussing the next steps necessary to take in norm building for AI policy.

1. ROLE OF NATIONS IN AI GOVERNANCE

Just as no nation is an island in cyberspace, so too may it be said that no nation can insulate itself from the myriad impacts of AI. Indeed,

---

[6] *Id*.
[7] *See* Eric Siegel, *The Media's Coverage of AI is Bogus*, SCI. AM. (Nov. 20, 2019), https://blogs.scientificamerican.com/observations/the-medias-coverage-of-ai-is-bogus/.
[8] *See* Urs Gasser & Carolyn Schmitt, *The Role of Professional Norms in AI Governance: Some Observations and Outline of a Framework*, MEDIUM (Apr. 25, 2019), https://medium.com/berkman-klein-center/the-role-of-professional-norms-in-ai-governance-some-observations-and-outline-of-a-framework-3dc25dcd2bdc.

nations—along with states, cities, and the private sector—are coming up with an array of strategies and principles on how best to harness the power of this coming wave to ensure that when it does wash ashore, economies, militaries, and societies are ready. This section focuses on the role of nations in AI governance at a macro-level, beginning by juxtaposing this topic with related debates on the role of the nation state in Internet governance and cybersecurity strategy in order to provide a foundation for comparative analysis.

1.1 Parallels with Internet Governance and Cybersecurity Strategies

As with AI governance, there has been increasing interest on the part of nations seeking to control cyberspace. A growing list of countries practice "cyber sovereignty" over their domestic Internet[9]; already, it has been reported that "two-thirds of all internet users [are] currently subjected to some degree of censorship of criticism aimed at the government, military, or ruling families."[10] Indeed, rather than degrading the idea of Westphalian sovereignty, in some ways cyberspace has given regimes around the world new tools to control restive populations through an array of cyber sovereignty campaigns, and applications with profound implications for human rights.[11] This wave of interest, which is being propounded by a range of authoritarian governments including China through its Belt and Road Initiative,[12] seek to "rewrite the rules of the Internet" by deepening and widening the role of nations in Internet governance, as compared to the big tent multi-stakeholder approach favored throughout the history of cyberspace.[13] These competing visions of Internet governance, including the extent to which nations will play a central or coordinating role, remain

[9] For an analysis of the Chinese approach to cyber sovereignty, see Scott J. Shackelford & Frank W. Alexander, *China's Cyber Sovereignty: Paper Tiger or Rising Dragon?,* POL'Y F. (Jan. 12, 2018), https://www.policyforum.net/chinas-cyber-sovereignty/.

[10] Andrea Little Limbago, *China's Global Charm Offensive*, WAR ON THE ROCKS (Aug. 28, 2017), https://warontherocks.com/2017/08/chinas-global-charm-offensive/.

[11] *See* EVGENY MOROZOV, THE NET DELUSION: THE DARK SIDE OF INTERNET FREEDOM 100 (2011). For more on this topic, see generally SCOTT J. SHACKELFORD, GOVERNING NEW FRONTIERS IN THE INFORMATION AGE: TOWARD CYBER PEACE (2020).

[12] *See* Samm Sacks, *Beijing Wants to Rewrite the Rules of the Internet*, ATLANTIC (June 18, 2018), https://www.theatlantic.com/international/archive/2018/06/zte-huawei-china-trump-trade-cyber/563033/.

[13] *See* Scott J. Shackelford & Amanda N. Craig, *Beyond the New 'Digital Divide': Analyzing the Evolving Role of Governments in Internet Governance and Enhancing Cybersecurity*, 50 STAN. J. INT'L L. 119, 120 (2014).

unresolved, although recent conferences since 2014 in Brazil, South Korea, and New York highlight ongoing support for the multi-stakeholder status quo.[14]

The race to develop effective AI strategies mirrors in many ways similar efforts to devise national cybersecurity strategies, which began in the early 2000s and have since rapidly picked up speed with more than seventy nations publishing such strategies as of April 2020.[15] Previous work has assessed how these strategies compare by considering their treatment of human rights,[16] along with cybercrime, critical infrastructure protection, and governance.[17] Among other things, this previous research has highlighted the extent to which nations are considering these issues through the lens of national security priorities, which may also be seen in the U.S. approach to AI strategy.

1.2 Assessing the U.S. National AI Strategy

The foundations of the United States' AI strategy largely took place in 2016 under the Obama administration when the White House launched a series of workshops and a subcommittee on Machine Learning and Artificial Intelligence.[18] These efforts led to the publication of three reports: The National Artificial Intelligence Research and Development Strategic Plan;[19] Artificial Intelligence, Automation, and the Economy;[20] and Preparing for the Future of Artificial Intelligence.[21] The former, *Preparing for the Future of Artificial Intelligence* surveyed the state of AI at the time, its potential applications and uses, explored

---

[14] *See id*.
[15] *See Cyber Security Strategies*, NATO CCDCOE, https://ccdcoe.org/library/strategy-and-governance/?category=cyber-security-strategies (last visited May 1, 2020).
[16] *See* Scott J. Shackelford, *Should Cybersecurity Be a Human Right? Exploring the 'Shared Responsibility' of Cyber Peace*, 55 STAN. J. INT'L L. 155, 156 (2019).
[17] *See* Scott J. Shackelford & Andraz Kastelic, *A State-Centric Cyber Peace? Analyzing the Current State and Impact of National Cybersecurity Strategies on Enhancing Global Cybersecurity*, 18 NYU J. OF LEG. & PUB. POL'Y 895 (2016)
[18] White House Press Release (Oct. 12, 2016), The Administration's Report on the Future of Artificial Intelligence, https://obamawhitehouse.archives.gov/blog/2016/10/12/administrations-report-future-artificial-intelligence.
[19] *See* Lynne E. Parker, *Creation of the National Artificial Intelligence Research and Development Strategic Plan*, 39 AI MAG. 39 (2018).
[20] *See* U.S. EXEC. OFF. OF THE PRES., ARTIFICIAL INTELLIGENCE, AUTOMATION, AND THE ECONOMY (2016), https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF.
[21] EXECUTIVE OFFICE OF THE PRESIDENT, PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE 48 (2016); Alan Bundy, Preparing for the Future of Artificial Intelligence 285-87, (2017).

multiple regulatory, policy, and governance issues related to AI, and made over 40 recommendations for future actions.[22]

The report titled *AI, Automation, and the Economy*,[23] was drafted as a follow up that broadly forecasts the effects of AI on the U.S. economy and workforce.[24] It makes recommendations in three broad categories: (1) maximizing the beneficial applications of AI; (2) training the workforce for jobs of the future; and (3) finding ways to assist workers in workforce transitions.[25] Relatedly, some have argued that the federal government still lacks a clear understanding of the capabilities of AI and its potential to affect various social and economic sectors including workforce impacts.[26]

Finally, the *National AI Research and Development Strategic Plan* has received the most attention of the three reports and largely constitutes the core of the current National AI Initiative.[27] The plan established seven broad priority areas for the U.S. government in relation to AI R&D.[28] It emphasized the role that the federal government plays in advancing research, development, and education activities in artificial intelligence through fostering coordination and collaboration between stakeholders to leverage intellectual, physical, and digital resources.[29]

In May 2018, the Trump Administration held a summit on Artificial Intelligence for American Industry[30] and put out a companion report emphasizing

---

[22] See *Preparing for the Future of Artificial Intelligence*, NAT'L SCI. & TECH. COUNCIL COMM. ON TECH. (2016). https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.

[23] ARTIFICIAL INTELLIGENCE, AUTOMATION, AND THE ECONOMY, *supra* note 20.

[24] *Id*.

[25] *Id*.

[26] *See* Justin Sherman, *Why the US Needs an AI Strategy,* WORLD POLITICS REV. (Mar. 14, 2019), https://www.worldpoliticsreview.com/articles/27642/why-the-u-s-needs-a-national-artificial-intelligence-strategy.

[27] *See also* P. Jonathan Phillips et al., *Four Principles of Explainable Artificial Intelligence*, NIST (2020), https://www.nist.gov/system/files/documents/2020/08/17/NIST%20Explainable%20AI%20Draft%20NISTIR8312%20%281%29.pdf (discussing explainable AI).

[28] See NAT'L SCI. & TECH. COUNCIL, NATIONAL AI RESEARCH AND DEVELOPMENT STRATEGIC PLAN (2016), https://www.nitrd.gov/pubs/national_ai_rd_strategic_plan.pdf (listing seven priority areas: "1) making long-term investments in AI research; 2) developing effective methods for human-AI collaboration; 3) understanding and addressing the ethical, legal, and societal implications of AI; 4) ensuring the safety and security of AI systems; 5) developing shared datasets and environments for AI training and testing; 6) measuring and evaluating AI technologies through standard benchmarks; and 7) better understanding national AI R&D workforce needs.").

[29] *Id*.

[30] *See* White House, *White House Hosts Summit on Artificial Intelligence for American Industry* (2018), https://www.whitehouse.gov/articles/white-house-hosts-summit-artificial-intelligence-american-industry/

the administration's support of private sector led development of AI technologies.[31] In the report, the White House announced plans to open data sources for private companies training AI technologies while also working within government to speed up the adoption of AI technologies in public services.[32] Soon after, the White House also released a short white paper and accompanying website titled *Artificial Intelligence for the American People* that outlined the Trump Administration's priorities for AI, which included: (1) Prioritizing funding for AI research, (2) removing regulatory barriers to the deployment of AI technologies, (3) training the future workforce, (4) achieving strategic military advantage, (5) leveraging AI government services, and (6) leading international AI negotiations.[33] These efforts were followed by an AI Summit in September 2019 that explored the use of AI in government.[34]

In February 2019, the Trump administration issued an executive order launching the American AI initiative.[35] Though the executive order and the initiative suggest that the federal government plays an important role in promoting AI Research and Development, the American AI initiative also calls for U.S. companies to "drive technological breakthroughs in AI across the Federal Government, industry, and academia in order to promote scientific discovery, economic competitiveness, and national security."[36] The American AI initiative was accompanied with a 2019 update of the previously published National AI Research and Development Strategic Plan.[37] The updated plan slightly adjusts the seven policy priorities of the Obama R&D strategy and adds public-private partnership as an eighth priority. This eighth priority calling for public-private partnerships continued the trend of the administration largely supporting

---

[31] *See Summary of the 2018 White House Summit on Artificial Intelligence for American Industry*, OFF. OF SCI. & TECH. POL'Y (May 10 2018), https://www.whitehouse.gov/wp-content/uploads/2018/05/Summary-Report-of-White-House-AI-Summit.pdf?latest.

[32] *Id*.

[33] See WHITE HOUSE ARTIFICIAL INTELLIGENCE FOR THE AMERICAN PEOPLE (May 10 2018), https://www.whitehouse.gov/briefings-statements/artificial-intelligence-american-people/

[34] *See Summary of the 2019 White House Summit on Artificial Intelligence for American Industry*, OFF. OF SCI. & TECH. POL'Y (Sept. 19, 2019), https://www.whitehouse.gov/wp-content/uploads/2019/09/Summary-of-White-House-Summit-on-AI-in-Government-September-2019.pdf

[35] *See* Exec. Off. of the Pres. Executive Order 13859: *Maintaining American Leadership in Artificial Intelligence* (Feb. 11 2019), https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence.

[36] *Id*.

[37] *See* NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN: 2019 UPDATE, NAT'L SCI. & TECH. COUNCIL (2019), https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf

development of AI being led by the private sector.[38] This is an important contrast when comparing the United States' strategic approach to AI compared to other countries (particularly China). [39]

In summary, when speaking of the U.S. AI strategy analysts are usually referring to the American AI initiative, updates to the AI R&D Strategic Plan, and the brief AI for the American People report.[40] The U.S. strategy, however, remains limited in scope and detail. Some timelines are given within the executive order discussed above,[41] but actual commitments are scarce, *i.e.* the strategy does not currently include any new funding.[42] Rather the initiative calls for the Office of Management and Budget to prioritize existing funding for AI research and President Trump's FY2020 budget included nearly $1 billion in funding for non-defense AI R&D. Furthermore, the President's budget request also called for increased investment in AI specifically, but decreases funding for federal research and development overall[43] The FY2021 budget request maintains the same trend in increasing funding for AI while decreasing R&D funding overall. There are concerns that other governments, particularly China, are far outspending the United States[44]. Government figures for China are hard to pin-point, but in comparison, the City Government of Shanghai alone plans to invest $15 billion on AI research and development over the next ten years.[45] In short, the U.S. AI Strategy, such as it is, in its current state lacks a clear timeline, measurable

---

[38] See Darell West, *Assessing Trump's Artificial Intelligence Executive Order*, BROOKINGS INST. (Feb. 12, 2019), https://www.brookings.edu/blog/techtank/2019/02/12/assessing-trumps-artificial-intelligence-executive-order/.

[39] See Kai-Fu Lee, *The Great AI Duopoly*, 36 NEW PERSPECTIVES Q. 27, 27 (2019).

[40] Despite the aforementioned diverging views on whether the United States has what can effectively be deemed a strategy or whether the United States is doing enough in this area, the National AI initiative refers to itself as the United States National strategy. *See Artificial Intelligence for the American People*, WHITE HOUSE (2021), https://www.whitehouse.gov/ai/executive-order-ai/ (last visited Jan 18, 2021).

[41]See Executive Order 13859 of Feb. 11, 2019, Maintaining American Leadership in Artificial Intelligence, 84 Fed. Reg. 3967–3972.

[42] *Id*.

[43] *See Analytical Perspectives*, OFF. OF MAN. & BUDGET (2019), https://www.whitehouse.gov/omb/analytical-perspectives/.

[44]see Thomas J. Colvin et al., *A Brief Examination of Chinese Government Expenditures on Artificial Intelligence R&D*, Sci. & Tech. Pol'y Inst. (2020), https://www.ida.org/-/media/feature/publications/a/ab/a-brief-examination-of-chinese-government-expenditures-on-artificial-intelligence-r-and-d/d-12068.ashx.

[45] *See* Daniel Ren, *Shanghai Aims to Raise US$15 Billion in Funds to Gain an Upper Hand in AI Development*, S. CHINA MORNING POST (July 5, 2018), https://www.scmp.com/business/companies/article/2153792/shanghai-aims-raise-us15b-funds-gain-upper-hand-ai-development.

milestones, or sufficient funding and thus provides fruitful ground for the incoming Biden administration for further engagement.

To fully assess the state of the U.S. AI strategy, however, given the degree of decentralized policymaking it is helpful to also consider developments in the U.S. Department of Defense (DoD), Congress, along with state and local governments.

1.2.1. Department of Defense Strategy and Investment in AI

The day after the release of the executive order discussed above, the U.S. Department of Defense (DoD) released its own AI strategy[46]. The DoD AI Strategy outlines four strategic focus areas: (1) delivering AI-enabled capabilities that address key missions; (2) partnering with leading private sector technology companies, academia, and global allies; (3) cultivating a leading AI workforce; and (4) leading in military ethics and AI safety.[47] Central to the DoD strategy is the establishment of the Joint Artificial Intelligence Committee (JAIC) "to accelerate the delivery of AI-enabled capabilities, scale the Department-wide impact of AI, and synchronize DoD AI activities to expand Joint Force advantages."[48] In 2018, DoD pledged $2 Billion through 2023 on AI research and development through the Defense Advance Research Project Agency (DARPA) in support of its AI Strategy.[49] These funds are in addition to ongoing R&D funding and does not include classified projects,[50] while the past several National Defense Authorization Acts (NDAs) have included multiple sections dealing directly with Artificial Intelligence.[51]

As required by the 2019 NDA, RAND was contracted as an independent auditor to assess the state of DoD's AI efforts and its ability to scale. RAND's report found that the DoD strategy lacked "baselines and metrics in conjunction with its AI vision" and that the JAIC lacked the visibility and authority to carry

---

[46] *See* Terri Moon Cronk, *DoD Unveils Its Artificial Intelligence Strategy*, DEFENSE.GOV (Feb. 12 2019), https://www.defense.gov/Explore/News/Article/Article/1755942/dod-unveils-its-artificial-intelligence-strategy/.

[47] *See* SUMMARY OF THE 2018 DEP. OF DEF. ARTIFICIAL INTELLIGENCE STRATEGY (2019), https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF.

[48] *Id*.

[49] *See DARPA Announces $2 Billion Campaign to Develop Next Wave of AI Technologies*, DARPA (Sept. 7, 2018), https://www.darpa.mil/news-events/2018-09-07.

[50] *Id*.

[51] *See* Cong. Gov. Legislation Search Results, https://www.congress.gov/search?searchResultViewType=expanded&q={%22source%22:%22legislation%22,%22search%22:%22Artificial+intelligence%22,%22bill-status%22:%22law%22}.

out its mission effectively.[52] In June 2020, following the RAND report, the DoD Inspector General also found that the DoD's strategic efforts were hindered by a lack of a clear organizational definition of AI, and thus lacked appropriate governance structures or consistent security controls.[53] In the future, there will likely need to be structural changes for the JAIC along the lines of the RAND report.[54] The JAIC will focus on DOD-wide AI transformation, moving away from building products.[55] The Defense Authorization Act also included funding for a congressional National AI initiative. Nowhere in these documents and initiatives, though, is the topic of human rights discussed, which is an omission that we unpack below.

1.2.2. U.S. Congress

In August 2018 Congress established the National Security Commission on Artificial Intelligence (NSCAI) as an independent bipartisan commission "to consider the methods and means necessary to advance the development of artificial intelligence, machine learning, and associated technologies to comprehensively address the national security and defense needs of the United States."[56] The Commission now releases quarterly recommendations to Congress.[57] The first set of recommendations were released in April 2020, which did not discuss human rights concerns other than noting a general need to "advance[e] ethical and responsible AI."[58] Senate Minority Leader Chuck Schuemer has also called for the creation of a new federal agency that would

---

[52] *See Department of Defense Posture for Artificial Intelligence: Assessment and Recommendations*, RAND CORP. (2019), https://www.rand.org/pubs/research_reports/RR4229.html.
[53] *See Audit of Governance and Protection of Department of Defense Artificial Intelligence Data and Technology*, INSPECTOR GENERAL DEP'T OF DEF. (June 29 2020), https://media.defense.gov/2020/Jul/01/2002347967/-1/-1/1/DODIG-2020-098.PDF.
[54] *See* Jackson Barnett, *Nearing Passage, the NDAA is Full of AI and Cyber Policy Changes*, FEDSCOOP (Dec. 9, 2020), https://www.fedscoop.com/ndaa-ai-house-law-cybersecurity-policy-changes/.
[55] *See* Jackson Barnett, *JAIC 2.0 Moves Away from Building Products to Focus on DOD-Wide AI Transformation*, FEDSCOOP (Nov. 6, 2020), https://www.fedscoop.com/jaic-2-0-moving-away-from-products-artificial-intelligence/.
[56] *See* Bert Chapman, *Literature Review: How US Government Documents Are Addressing the Increasing National Security Implications of Artificial Intelligence*, J. ADVANCED MILITARY STUD. (2020).
[57] *Id*.
[58] *See* NATIONAL SECURITY COMM'N ON ARTIFICIAL INTELLIGENCE, KEY CONSIDERATIONS FOR RESPONSIBLE DEVELOPMENT AND FIELDING OF ARTIFICIAL INTELLIGENCE (2020), *available at* https://www.nscai.gov/reports.

invest an additional $100 billion over five years on basic research in AI, but as of this writing the legislation has not advanced.[59]

The Congressional Artificial Intelligence Caucus was launched in 2017.[60] It is currently chaired by Pete Olson (R, TX-22) and Jerry McNerney (D, CA-09).[61] Its members have supported and introduced multiple house bills on AI.[62] The Senate Artificial Intelligence Caucus was launched in March 2019 after the introduction of the Trump administration's National AI Initiative.[63] The bipartisan caucus is led by Senators Martin Heinrich (D-N.M.) and Rob Portman (R-Ohio).[64] The National Artificial Intelligence Initiative Act and the National AI Research Resource Task Force Act were both included in the FY2021 Defense Authorization Act, which increased funding for AI R&D and use across government.

1.2.3 State and Local Developments

Multiple states have created AI taskforces including New York,[65] Vermont,[66] and Washington.[67] At a city level, many municipalities have incorporated AI into their smart city strategies and plans. Stockton California was the first city in the United States to release a strategy specifically focused on AI and the future of work and is currently running a well-publicized Universal Basic Income Trial for 500 residents as one potential policy response to the disruptions

---

[59] *See* Sebastian Moss, *Senator Schumer Proposes New US Government Agency With $100bn AI Budget*, DCD (Nov. 6, 2019), https://www.datacenterdynamics.com/en/news/senator-schumer-proposes-new-us-government-agency-100bn-ai-budget/.

[60] *See Delaney Launches Bipartisan Artificial Intelligence (AI) Caucus for 115th Congress*, Cong. A.I. Caucus (May 24, 2017), https://artificialintelligencecaucus-olson.house.gov/media-center/press-releases/delaney-launches-ai-caucus.

[61] *Id*.

[62] *See, e.g.*, H.R. 6216, 116 Cong. (2020).; FUTURE of Artificial Intelligence Act of 2020, H.R. 7559, 116 Cong. (2020).; AI Use in Government Act, H.R. 2575, 116 Cong. (2019).

[63] See *Portman, Heinrich Launch Bipartisan Artificial Intelligence Caucus,* Rob Portman (Mar. 13, 2019), https://www.portman.senate.gov/newsroom/press-releases/portman-heinrich-launch-bipartisan-artificial-intelligence-caucus.

[64] *Id.*

[65] *See* Albert F. Cahn, *The First Effort to Regulate AI was a Spectacular Failure*, FastCompany (Nov 26. 2019), https://www.fastcompany.com/90436012/the-first-effort-to-regulate-ai-was-a-spectacular-failure.

[66] *See* Grace Elleston, *As Artificial Intelligence Grows in Vermont, task Force Mulls State Policies,* VTDigger (Nov. 10, 2019), https://vtdigger.org/2019/11/10/as-artificial-intelligence-grows-in-vermont-task-force-mulls-state-policies/.

[67] *See* S.B. 5527, H.B. 1655, 66th Leg., 2019 Reg. Sess. (Wash. 2019)

being caused by AI.[68] Studies have indicated that AI-driven job displacement will have uneven effects on the U.S. workforce and estimate that job displacements will be as high as sixty-four percent in some municipalities.[69] A handful of AI-related bills have similarly been introduced at state and local levels. California has been the most proactive U.S. state by far in this vein, as seen in the 2018 Consumer Privacy Act (CPA) and its adoption of the Asilomar AI principles.[70]

1.2.4 International Collaboration

The federal U.S. AI strategy, such as it is, calls for maintaining U.S. leadership in AI while increasing international collaboration. Though the authors of this piece acknowledge the prevalence of an "AI race" narrative, we think it important to consider the negative implications of such rhetoric. Cave and ÓhÉigeartaigh, for example have argued that the "AI race" narrative presents a multitude of risks including incentives to "cut corners" around AI ethics and safety.[71] The AI-race narrative encourages competition, which may make international coordination and collaboration on norms and governance more difficult. In Parts 3 and 5 we make a broader call for polycentric approaches to AI governance to better conceptualize the distributed governance structure this domain. This study highlights areas of convergence in international strategy documents that point toward the possibility of fostering collaborative efforts around AI governance and policymaking. Part 2 analyzes existing national AI strategies and demonstrates multiple areas of potential collaboration that the U.S. government, particularly the Biden administration, could pursue in fostering international collaborative efforts.

In June 2020, together with Australia, Canada, France, Germany, India, Italy, Japan, Mexico, New Zealand, the Republic of Korea, Singapore, Slovenia, the United Kingdom, and the European Union, the United States launched the Global Partnership on Artificial Intelligence (GPAI).[72] The GPAI builds from the OECD AI recommendations that the U.S. signed and adopted in May 2019.[73] This development points to the desire among governments to collaborate around AI policy and governance, which likewise points to the contribution of the current

---

[68] See Hannah Miller & Isak Nti Asare, *Why Every City Needs to Take Action on AI*, OXFORD INSIGHTS (Aug. 1, 2018), https://www.oxfordinsights.com/insights/2018/8/1/why-every-city-needs-to-take-action-on-ai.
[69] *Id*.
[70] *See State of California Endorses Asilomar AI Principles*, FUTURE OF LIFE INST. (Aug. 31, 2018), https://futureoflife.org/2018/08/31/state-of-california-endorses-asilomar-ai-principles/.
[71] *See* Seán S. ÓhÉigeartaigh, *An AI race for Strategic Advantage: Rhetoric and Risks*, *in* PROC. OF THE 2018 AAAI/ACM CONF. ON AI, ETHICS, AND SOCIETY 36-40 (2018).
[72] *See Joint Statement from Founding Members on the Global Partnership on Artificial Intelligence*, U.S. DEP'T OF STATE (June 15, 2020), https://www.state.gov/joint-statement-from-founding-members-of-the-global-partnership-on-artificial-intelligence/.
[73] *See AI Principles*, OECD (2019), https://www.oecd.org/going-digital/ai/principles/.

Article. Identifying opportunities for norm creation based on current national AI initiatives is a necessary first step in establishing lasting international collaborative efforts to ensure the benefits of AI, including its impacts on human rights.

2. ANALYSIS OF AI STRATEGIES

This Part analyzes AI strategies to highlight areas of policy convergence and divergence that could lead to opportunities for norm development, and eventually customary international law. We begin by discussing the methodology utilized in this study before moving on to analyze the dimensions surveyed.

2.1 Methodology

This Article makes an original contribution by conducting content analysis on more than forty existing national AI strategies. To limit the scope of this study, we only analyzed national initiatives on AI similar to the U.S. case study in Part 1. Appendix 1 contains a list of the documents surveyed. Cross-national or regional strategies (e.g. the European Union's AI Strategy) were excluded. Likewise, countries that have initiatives but no document at the time of publication were excluded. The UAE for example, was the first country to appoint a Minister for Artificial Intelligence in government and has a website on its strategy,[74] but the policy document could not be found; it was therefore, excluded from the analysis. Similarly, examples such as Kenya, which has established an AI taskforce but has yet to publish their findings, were likewise excluded. Countries with multiple national AI initiatives, policies, or strategy documents only had one representative publication included in the parsing and quantitative analysis so as not to skew the results.[75]

[74] See *The UAE Seeks to be a Major Hub for Developing AI Techniques and Legislation*, NAT'L PROG. ARTIFICIAL INTELLIGENCE, https://ai.gov.ae/ (last visited Jan. 5, 2021).

[75] In conducting the qualitative content analysis, we found that many of these additional texts were specialized in nature, for example focusing specifically on the future of work, ethics, or research development. We felt that their inclusion in the word parsing would skew the results and as such opted to exclude them, with the exception of the Consultation on the OPC's Proposals for ensuring appropriate regulation of artificial intelligence, which specifically examines Canada's approach to AI privacy laws. In total, twenty-seven national initiatives were analyzed using our word parser

We conducted both quantitative and qualitative comparative content analysis of each document. Content analysis is a flexible research methodology that has been widely used in several disciplines. Klaus Krippendorff defines content analysis as "a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the context of their use."[76] This method "uses analytical constructs, or rules of inference, to move from the text to the answers to the research questions."[77] To this end, in our quantitative analysis we equated the number of times that keywords were mentioned in the document as a rough proxy for the focus of the strategy.

This quantitative analysis allows for reproduction using different themes, keywords, or documents. It also permits the longitudinal study of emerging trends as new initiatives are published or old initiatives changed. We selected the dimensions to be surveyed based on existing analyses of AI initiatives. We then built word lists based on existing literature on each theme. These word lists are included at the end of each subsection. In addition to this, Appendix 2 contains graphs reflecting how much each dimension is represented in each country surveyed, broken down by the number of times categorical keywords appear. Appendix 3 contains a breakdown of how often each specific keyword appears across the strategies and documents examined, organized by dimension.

Using these terms, we used a word parser to determine the relative percentages of each theme within each document. We have tracked the total word count of each term or set of terms as they appear in the collection of policies we studied to see if our word choice was effective. Finally, in order to study the level of convergence between different countries, we took the Standard Deviation of each dimension to see if the averages show significant similarities in the policy sets. This quantitative content analysis was supplemented by qualitative content analysis focusing on meaning, intentions, and policy context.

2.2 Dimensions Surveyed

[76] James Gunthrie et al., *Using Content Analysis as a Research Method to Inquire into Intellectual Capital Reporting*, 5 J. INTELLECTUAL CAPITAL 282, 290 (2004).
[77] JULIAN LABOY, FROM TAO TO PSYCHOLOGY: AN INTRODUCTION TO THE BRIDGE BETWEEN EAST AND WEST 28 (2012).

This section summarizes the key findings across the seven dimensions surveyed in this study, including: transparency, accountability, security, privacy, fairness, human-centered design, and public benefit. Many of these dimensions include significant overlap with human rights issues and concerns,[78] as is discussed below.

2.2.1 Transparency

Transparency, which may be defined as "openness; clarity; lack of guile and attempts to hide damaging information,"[79] is commonly used in the context of financial disclosures and/or organizational policies and practices. When used in the context of AI, transparency has come to signify openness as it relates to a particular aspect of the AI, for example transparency into the inner workings of artificial intelligence models.[80] This transparency would ideally not only be apparent to system engineers, but also be conveyed in such a way that all humans (including consumers) interacting with the AI will understand how information is used and how decisions are made.

Transparent AI is often made analogous with explainable AI,[81] trustworthy AI,[82] responsible AI,[83] and vice-versa.[84] When discussing transparency in AI, scholars, analysts, and commentators often describe AI

---

[78] *See, e.g.*, Scott J. Shackelford, *Should Cybersecurity Be a Human Right? Exploring the 'Shared Responsibility' of Cyber Peace*, 55 STAN. J. INT'L L. 155 (2019).

[79] InterPARES Trust, https://interparestrust.org/terminology/term/transparency (last visited Jan. 5, 2021).

[80] *See* Catherine Yeo, *What is Transparency in AI?*, FAIR BYTES (May 20, 2020), https://medium.com/fair-bytes/what-is-transparency-in-ai-bd08b2e901ac.

[81] A 2020 report by Deloitte said that "Transparent AI is explainable AI. It allows humans to see whether the models have been thoroughly tested and make sense, and that they can understand why particular decisions are made" See Deloitte (2020) "Transparency and Responsibility in Artificial Intelligence: a call for explainable AI" pg. 6 https://www2.deloitte.com/xe/en/insights/industry/dcom/a-call-for-transparency-and-responsibility-in-artificial-intelligence.html

[82] *See* Irfan Saif & Beena Ammanath, *Trustworthy AI is a Framework to Help Manage Unique Risk,* MIT TECH. REV. (2020), https://www.technologyreview.com/2020/03/25/950291/trustworthy-ai-is-a-framework-to-help-manage-unique-risk/#:~:text=For%20AI%20to%20be%20trustworthy,decisions%20must%20be%20fully%20explainable.

[83] See *Responsible AI: A Framework for Building Trust in Your AI Solutions*, ACCENTURE (2018) https://www.accenture.com/_acnmedia/PDF-92/Accenture-AFS-Responsible-AI.pdf.

[84] *See, e.g.*, EUR. COMM'N, ETHICS GUIDELINES FOR TRUSTWORTHY AI (2019), https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

as a ”black-box” due to the opaque or closed nature of many AI systems.[85] There is also some discussion in the literature on using transparent AI to counter data and algorithmic bias.[86] As seen in the Appendices, each of these terms and themes were captured in our word list.[87] Figure 1 highlights the prevalence across all documents of each of the terms reviewed.

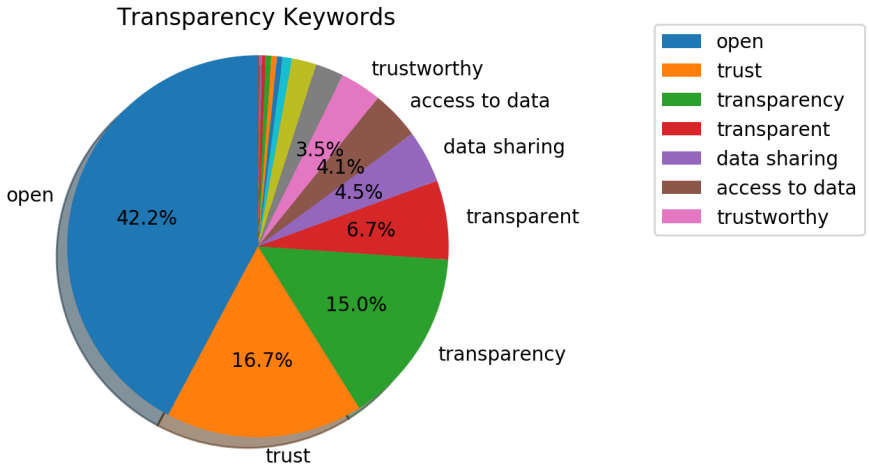

Figure 1: *Transparency Keyword Representation*

We found that many of the strategies contain explicit statements on transparency. Uruguay, for example states that “AI solutions used in the public sphere must be transparent […]” and that this transparency must: “[m]ake available the algorithms and data used for training the solution and its implementation, as well as the tests and validations performed [and] explicitly make visible, through active transparency mechanisms, all those processes that use AI.”[88] As this example indicates, much of the emphasis on transparency in these strategies focuses on the use of AI in government and in administering public services. Norway’s AI Strategy, for example, states explicitly that the government will “set requirements

[85] See Will Knight, *The Financial World Wants to Open AI’s Black Boxes*, MIT Tech. Rev. (Apr. 13, 2017), https://www.technologyreview.com/s/604122/the-financial-world-wants-to-open-ais-black-boxes..

[86] *See* G.S., Nelson, *Bias in Artificial Intelligence.* 80 N.C. Medical J. 220-22 (1998).

[87] ’Responsible AI‘ and its cognates were removed from the word list due to the prevalence of false positives in the initial parsing. It can be argued that transparency overlaps and intersects with accountability which we have chosen to treat as a separate theme in our analysis.

[88] Uruguay - Agencia de Gobierno Electrónico y Sociedad de la Información y Conocimiento, Artificial Intelligence Strategy for the Digital Government 9 (2019).

for transparency and accountability in new public administration systems in which AI is part of the solution."[89] Italy's statement on the topic of transparency is similar.[90]

Several countries emphasized the need for further R&D in creating transparent AI systems. The U.S. Strategy, for example, states: "To garner trust and confidence, AI technologies should be transparent in how they work and provide reasonable guarantees on the safety, security, robustness, and resiliency of their operation. Many existing AI systems, however, lack these characteristics due to unsolved technical hurdles that require further R&D."[91] Likewise, Lithuania's Strategy maintains that "AI applications should be ethical, safe, reliable and transparent."[92] Though the principles of transparency and trustworthiness are espoused in many of the national initiatives, Finland notes that "it has yet to be specified what these principles mean in practice from the viewpoint of various actors and regulatory systems"[93] this presents a clear opportunity for international collaboration to define and establish clear frameworks for transparent AI. This will almost certainly require an emphasis on public-private partnerships, which should deepen the opportunities and incentives for international cooperation.

There is a tension, particularly among those countries that position themselves primarily as users of AI technology (rather than creators of new solutions), between transparency and the use of proprietary solutions. Many strategies discuss the need to establish standards, guidelines, and procedures for algorithmic transparency. It is unclear how this could be done effectively without cross border collaboration. India emphasizes this point in saying that "Opening the Black Box, assuming it is possible and useful at this stage, should not aim towards opening of code or technical disclosure – few clients of AI solutions would be sophisticated AI experts

---

[89] Norwegian Ministry of Local Government and Modernisation, *National Strategy for Artificial Intelligence* (2020), https://www.regjeringen.no/contentassets/1febbbb2c4fd4b7d92c67ddd353b6ae8/en-gb/pdfs/ki-strategi_en.pdf (last visited Jan 18, 2021).

[90] The Agency For Digital Italy, *Artificial Intelligence at the Service of Citizens* (2018), https://ia.italia.it/assets/whitepaper.pdf (last visited Jan 18, 2021) ("The issue of the responsibility of public administration also has to do with the duties of the latter with respect to citizens, when it decides to provide them with services or to make decisions that concern them, using Artificial Intelligence solutions. The functioning of the latter must meet criteria of transparency and openness. Transparency becomes a fundamental prerequisite to avoid discrimination and solve the problem of information asymmetry, guaranteeing citizens the right to understand public decisions.").

[91] Nat'l Sci. Tech. Council, *supra* note 28.

[92] *Id*.

[93] *Id*.

- but should rather aim at "explainability."[94] With extended disclosure, though, what needs to be balanced is whether the algorithm's parameter may induce the individuals and companies to change their behavior and in turn game the system. Clearly, more collaborative research is required in this area."[95]

transparency	transparent
open	openess
closed	trustworthy
trust	explanability
explainable	understandable
black-box	black box
opacity	opaque
data bias	available data
algorithmic bias	data sharing
loss of trust	access to data

Table 1: *Transparency Keywords*

2.2.2 Accountability

Accountability, which may be understood as "responsible or answerable,"[96] in the context of AI is often considered as being responsible for what you do and are able to give reasons for the actions, or choice. Within AI, the topic of accountability is one of considerable debate due to the distance between the initial programming and the outcome, leading to questions about who is liable for the actions made by AI. In our definition of accountability, we have examined the need for accountability in the design process, within deployment, and after any sort of harm has occurred. Thus, we have focused on terminology such as 'oversight' and 'framework' to address the need for accountability in AI development, and 'liability' and 'victim' to cover the need for redress in the event of a failure. In this dimension, it is important to note that while there is a general convergence on the amount of time accountability was discussed, this does not indicate that there was a consensus on how accountability would be best achieved. Some documents will include references to specific entities that would be liable if harm occurs from AI.

[94] *Pan-Canadian AI Strategy,* CIFAR (2017), https://cifar.ca/ai/ (last visited Jan 18, 2021).
[95] NITI AAYOG, NATIONAL STRATEGY FOR ARTIFICIAL INTELLIGENCE 86 (2018).
[96] Definition of "Accountability," https://www.dictionary.com/browse/accountable (last visited Jan. 25, 2021).

During our research, we identified several general terms such as 'framework', 'model', and 'law' which have been broadly applied in the policies. Unsurprisingly, given that the documents are centered around policy, these keywords appear quite frequently without substance, being sprinkled in with generic qualifiers such as 'ethical AI.' The most telling discussions around accountability frequently accompanied the 'responsibility' keyword, implicating what principles were important to develop frameworks, governance, and regulations around.

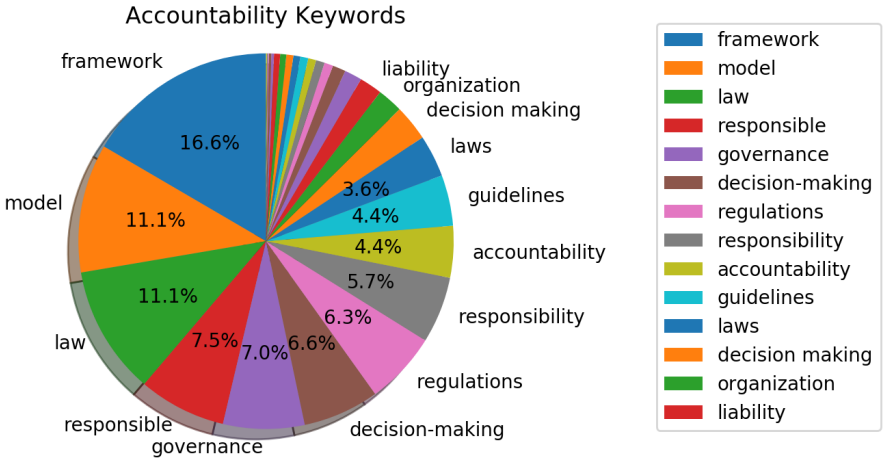

Figure 2: ***Accountability Keyword Representation***

Nations seem to be converging upon the idea that in order to utilize AI, it must be done within a responsible framework. Some countries set a timeline of when they want a regulatory system in place to govern AI, such as Russia, which plans to have a flexible regulatory system in place by 2030.[97] Denmark's policy specifies that they will have a working group to examine and apply existing law to AI, and if there are gaps present there "may be a need to launch legislative initiatives at national or EU level."[98] A microcosm which exposes the need for specialized AI regulation is that of autonomous vehicles (AVs). If an autonomous vehicle is involved in an accident, then it is often unclear on to whom the blame

[97] Decree of the President of the Russian Federation—On the Development of Artificial Intelligence in the Russian Federation, CSET, at 17 (Oct. 28, 2019), https://cset.georgetown.edu/research/decree-of-the-president-of-the-russian-federation-on-the-development-of-artificial-intelligence-in-the-russian-federation/.

[98] *See Danish Government National Strategy for Artificial Intelligence*, Ministry of Fin. & Ministry of Indus. Bus. & Fin. Affs. (Mar. 2019), https://en.digst.dk/media/19337/305755_gb_version_final-a.pdf.

lies – the owner of the vehicle, the company that developed the software, or perhaps even the car manufacturer. New Zealand's statement on AI discusses this dilemma.[99] While there are no definite answers at this given moment, New Zealand's Allen Institute for Artificial Technology have created three rules for regulating AI:

> *1. An AI system must be subject to the full gamut of laws that apply to its human operator.*
> *2. An AI system must clearly disclose that it is not human.*
> *3. An AI system cannot retain or disclose confidential information without explicit approval from the source of that information.*[100]

Even if these regulations were to be put into place, there remains the question of oversight. China's *A Next Generation Artificial Intelligence Development Plan* proposes an AI supervision system, with a two-tiered structure to manage the entire process of AI development, from design to result application.[101] This supervision system would encourage "AI industry and enterprise self-discipline, and earnestly strengthen management, increase disciplinary efforts aimed at the abuse of data, violations of personal privacy, and actions contrary to moral ethics,"[102] suggesting this system would be implemented in all private sector companies that create and use AI. Supervision is also a necessary element in accountability because AI may also make mistakes. For example, in the healthcare context, it is imperative that a doctor does not settle for a diagnosis or treatment plan simply because it was suggested by the AI, especially when a better alternative may exist.[103]

While oversight and regulations may provide a beginning to AI legal policy, there must also be dialogue between the government, corporations, academia, and civil society to identify any potential accountability gaps. Canada's policy encourages such active discourse

---

[99] *Artificial Intelligence: Shaping a Future New Zealand*, A.I. F. N. Z., (May 2018), https://www.mbie.govt.nz/dmsdocument/5754-artificial-intelligence-shaping-a-future-new-zealand-pdf; Scott J. Shackelford & Rachel Dockery, *Governing AI*, __ CORNELL J. L. PUB. POL'Y __ (forthcoming 2021) (discussing the prospects of governing AI through a polycentric framework with an AV case study).
[100] Oren Etzioni, *How to Regulate Artificial Intelligence,* N.Y. TIMES (Sept. 1, 2017), https://www.nytimes.com/2017/09/01/opinion/artificial-intelligence-regulations-rules.html.
[101] *Next Generation Artificial Intelligence Development Plan*, CHINA SCI. & TECH. NEWSL. (Sept. 15, 2017), http://fi.china-embassy.org/eng/kxjs/P020171025789108009001.pdf.
[102] *Id*.
[103] *AI White Paper*, AGENZIA PER L'ITALIA DIGITALE, at 16 (May 4, 2018), https://ia.italia.it/en/assets/whitepaper.pdf.

between these separate groups to "ensure that harms are identified and addressed, and that policy adequately reflects public interest objectives and addresses concerns from specific groups."[104] Suggested strategies include running hackathons, public consultations, and increasing engagement with stakeholders when developing policies. Denmark will label brands and products that utilize ethical data practices to encourage accountability within the business sector.[105] Discussions on accountability within the private sector have also included equipping the workforce with the right skillset to be able to integrate AI into their workflow. Finland emphasizes the idea that AI technologies will result in job losses, and that it is society's responsibility to take responsibility for these losses. As AI technology was funded primarily by taxpayer funds, Finnish leaders argue that it is unfair for citizens to be negatively affected by these very capabilities; a potentially potent line of argument for other nations seeking to safeguard human rights in the AI Age.[106] In addition to this line of argument, Italy's white paper presents the idea that the States bears a responsibility to create an educational system that will keep up with the changing landscape shaped by AI.[107] While the State bears responsibility, it must also collaborate with academia to ensure that there are enough AI professionals which have the necessary skillset to develop and effectively utilize AI technologies in a capable, ethical manner.

As is apparent, accountability is a quite broad topic in AI, and there are several more factors that remain unexamined in this section. One of these is accountability within data governance, which places a responsibility on the State to protect public data from misuse.[108] This topic will be further covered in our section about privacy, where data governance will be addressed. Additionally, there also exists a responsibility to protect vulnerable populations from discrimination caused or exacerbated by AI technologies. This will be covered in our section on Fairness below, which specifically concerns topics on equality, bias, and equity.

---

[104] *Rebooting Regulation: Exploring the Future of AI Policy in Canada*, CIFAR, at 7 (May 2019), https://cifar.ca/wp-content/uploads/2020/01/rebooting-regulation-exploring-the-future-of-ai-policy-in-canada.pdf.

[105] *The Danish Government National Strategy for Artificial Intelligence*, Ministry of Fin. & Ministry of Indus. Bus. & Fin. Affs., at 31 (Mar. 2019), https://en.digst.dk/media/19337/305755_gb_version_final-a.pdf.

[106] *Id*. at 51.

[107] *See* Agenzia per l'Italia Digitale, *supra* note 103.

[108] *Data Protection Law: An Overview*, Cong. Rsch. Serv. (Mar. 25, 2019), https://fas.org/sgp/crs/misc/R45631.pdf.

decision making	impact assessment
risk management	internal control
accountability	external control
accountable	responsible
responsibility	justice
regulate	law
regulations	liability
liable	governance
govern	causality
compensate	compensation
victim	victims
laws	decision-making
guidelines	oversight
audit	auditing
redress	sandbox
framework	organization
decentralized	model

Table 2, *Accountability Keywords*

Overall, there has not been a nation-state level, comprehensive regulatory framework developed to govern AI. Many countries discuss the possibility of implementing such a regime, but these plans remain nascent as of this writing.. There is likewise divergence on the extent that the State, the private sector, and/or civil society bears responsibility for the myriad effects AI, a similar debate that is playing out in discussing AI security.

2.2.3 Security

In general, security may be defined as being free from danger.[109] However, in the context of technology, security is often thought of in terms of cybersecurity, which is defined by the U.S. Cybersecurity and Infrastructure Security Agency (CISA) as "art of protecting networks, devices, and data from unauthorized access or criminal use and the practice of ensuring confidentiality, integrity, and availability of information."[110] Typically, these definitions will address the security of the AI from external threats, although security also plays a role in internal threats such as fail-safe mechanisms. Within AI security, the

[109] Definition of 'Security,' https://www.merriam-webster.com/dictionary/security (last visited Jan. 6, 2021).
[110] *What is Cybersecurity?*, https://us-cert.cisa.gov/ncas/tips/ST04-001 (last visited Jan. 6, 2021).

overall safety of the technology involves leveraging AI to identify and mitigate cyber threats with less human intervention than is typically expected. Within the keywords selected, the dimension of 'security' will cover a broad range of topics, from implementing security features into AI to utilizing AI to achieve security goals.

The most common keyword in the documents surveyed was 'security' by a wide margin. Nevertheless, it must be noted that 'security' sometimes appeared with other terms, such as 'social security' or 'job security', which did not necessarily represent the dimension being examined. Overall, the discussion of security encompasses a wide range of considerations, many of which are addressed in this section. Our keywords also took into account basic cybersecurity principles, such as availability, integrity, and confidentiality of data, but these were not widely discussed in the policy papers surveyed in this study.

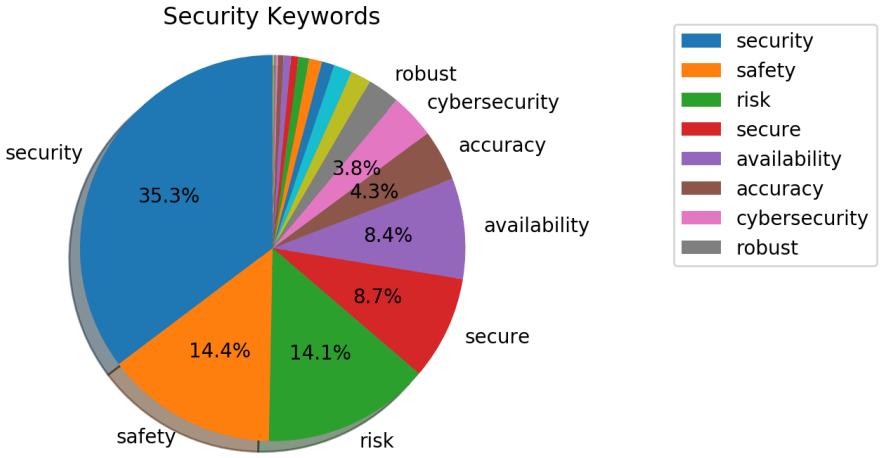

Figure 3: *Security Keyword Representation*

The international community acknowledges that security is an important consideration when formulating an AI strategy, but there is a strong divergence in terms of where and how this security should be applied. In U.S. AI policy, there is a strong emphasis on employing AI to enhance national security, both by developing offensive AI technology and also by defending against attacks driven by AI.[111] DARPA and Intelligence Advanced Research Projects Activity (IARPA) have created a variety of programs designed to combat attacks against AI, such as Secure, Assured, Intelligent Learning Systems (SAILS), Trojans in Artificial Intelligence (TrojAI), and Guaranteeing AI Robustness against

[111] *See* 2019 UPDATE, *supra* note 37.

Deception (GARD).[112] The United Kingdom takes a similar approach to the United States, having created the National Security Strategic Investment Fund, which would contribute up to £85m in advanced technologies to protect national security.[113] New Zealand's policy specifically mentions the need to bolster political security against AI related attacks such as social media manipulation and the related issue of deep fakes.[114]

Discussions on national security are sometimes supplemented by considerations of implementing AI security in the private sector. Critical infrastructure, which is often managed by the private sector,[115] is particularly vulnerable to cyber attacks, and the growth of AI technology will only exacerbate these risks. India presents a policy whereby the private sector will be held accountable for AI security. The policy revolves around negligence, where the neglect of security could result in serious fines for the company, but the law introduces safe harbors for companies who take appropriate steps to monitor, test, and improve AI products.[116] Similar safe harbor laws focusing on cybersecurity have been passed by a variety of U.S. states, including Ohio.[117]

One aspect of AI security that many nations seem to be converging on is the need to implement security into the basis of the AI's design. The United States, Sweden, France, and Germany mention that "safety and security considerations cannot be an afterthought; they must be an integral part of the early design stage."[118] South Korea takes a proactive approach to AI security by creating a policy to implement quantum computing to reduce the risk of cyber-attacks at their root. By 2020, South Korea vowed to test quantum cryptography on exclusive networks to maximize security for national facilities.[119] In 2025,

---

[112] *Id*. at 23.
[113] *Industrial Strategy: Building a Britain fit for the future*, HM Gov't (2017), at 180, http://www.gov.uk/beis.
[114] *See Artificial Intelligence: Shaping a Future New Zealand*, A.I. F. N. Z., (May 2018), https://www.mbie.govt.nz/dmsdocument/5754-artificial-intelligence-shaping-a-future-new-zealand-pdf
[115] *Critical Infrastructure Sectors*, CISA, https://www.cisa.gov/critical-infrastructure-sectors (last visited Jan. 6, 2021).
[116] *See National Strategy for Artificial Intelligence*, NITI Aayog (June 2018).
[117] *See* Michael Kassner, *Ohio Law Creates Cybersecurity 'Safe Harbor' for Businesses*, Tech. Rep. (Jan. 3, 2019), https://www.techrepublic.com/article/ohio-law-creates-cybersecurity-safe-harbor-for-businesses/.
[118] *See National Approach to Artificial Intelligence*, Gov't Off. of Swed. (2018), at 5, http://www.government.se/; *French Strategy for Artificial Intelligence*, AI for Humanity (Mar. 29, 2018), https://www.aiforhumanity.fr/en/.
[119] *See Mid-to Long-Term Master Plan in Preparation for the Intelligent Information Society*, Gov't of the Republic of Korea (2018), at 39, http://www.msip.go.kr/dynamic/file/afieldfile/msse56/1352869/2017/07/20/Master%20Plan%20for%20the%20intelligent%20information%20society.pdf..

these quantum code protected networks would include other sectors, such as healthcare and finance, and by 2030 South Korea intends to develop the core technology for a quantum Internet.[120]

Most of the discussion until this point has centered upon securing AI technologies from outside intervention, but it is important to note that AI can also be used to increase security. The UK government, for example, has argued that "[m]achine learning can identify, categorize and analyze these more effectively than individual researchers. By working simultaneously on different tasks, across a large number of devices and systems, AI can help defend against large attacks."[121] While attackers will be utilizing AI to target cyber vulnerabilities, AI can also be employed by defenders. Robots also have the capacity to assist people in natural disasters, where traditional rescue crews may not be able to function. Germany, for example, has "plans for robots to be used especially in critical circumstances arising in an inhospitable environment, for instance when there has been a calamity in a chemical factory or when the structure of buildings has to be assessed in the wake of an earthquake."[122]

secure	security
confidentiality	availability
integrity	vulnerability
vulnerabilities	vulnerable
compromise	safety
reliability	robust
robustness	predictability
repeatability	accuracy
reproducibility	cybersecurity
data management	

Table 3, *Security Keywords*, September 2020

Finally, there have been discussions on using AI to improve policing and security via surveillance.[123] Such efforts, though, are double-edged and can

[120] *Id*. at 39.
[121] *See* Dame W. Hall & Jérôme Pesenti, *Growing the Artificial Intelligence Industry in the UK*, at 21 (Oct. 15, 2017), https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk.
[122] *See Artificial Intelligence Strategy*, Nationale Strategie fur Kunstliche Intelligenz, at 17 (Nov. 2018), http://www.ki-strategie-deutschland.de/.
[123] *See Mid-to Long-Term Master Plan in Preparation for the Intelligent Information Society*, Gov't of the Republic of Korea, at 13 (2018), http://www.msip.go.kr/dynamic/file/afieldfile/msse56/1352869/2017/07/20/Master%20Plan%20for%20the%20intelligent%20information%20society.pdf

promote both public safety along with enabling authoritarian regimes. China, for example, is pushing for the development of new AI technologies to assist law enforcement, such as "detection sensor technology, video image information analysis and identification technology, biometric identification technology, intelligent security and police products."[124] Nonetheless, these investments have been noted to carry the risk of mass surveillance, human rights violations, and reduce individual autonomy,[125] which is the next topic of discussion.

2.2.4 Privacy

Privacy was famously defined as "the right to be let alone"[126] in the nineteenth century, but by the twenty-first century privacy has become a vast concept encompassing (among much else) freedom of thought, bodily integrity, solitude, information integrity, freedom from surveillance, along with the protection of reputation, and personality.[127] Still, there are widely differing views as to the bounds of privacy rights, including whether it should be considered a property right (and whether companies like Facebook, for example, should pay users for their information),[128] and especially how to update core privacy concepts for the AI Age.

We have found that the term which appears the most from this category is 'private.' Upon examination of the source material, however, we have found that most of the word usage revolves around the 'private sector,' or 'private companies,' rather than private information. The keywords 'privacy' and 'personal' were usually accompanied by rather insightful discussions on the matter. Many nations present privacy as a fundamental consideration when shaping AI policy, but there is a wide divergence on how this complex concept should be implemented. For example, some AI applications require an enormous amount of data, which itself has privacy and human rights implications,[129]

---

[124] *See Next Generation Artificial Intelligence Development Plan*, China Sci. & Tech. Newsl., at 20 (Sept. 15, 2017), http://fi.china-embassy.org/eng/kxjs/P020171025789108009001.pdf.
[125] *See French Strategy for Artificial Intelligence*, AI for Humanity , at 124 (Mar. 29, 2018), https://www.aiforhumanity.fr/en/.
[126] Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 Harv. L. Rev. 193, 195 n.4 (1890); Roe v. Wade, 410 U.S. 113, 152 (1973).
[127] *See generally* Daniel J. Solove, *Conceptualizing Privacy*, 90 Calif. L. Rev. 1087 (2002) (advocating a pragmatic approach to conceptualizing privacy); *Fragile Merchandise: A Comparative Analysis of the Privacy Rights of Public Figures*, 19 Am. Bus. L.J.125, 125-27 (2012).
[128] *See, e.g.*, Leonid Bershidsky, *Let Users Sell Their Data to Facebook*, Bloomberg (Jan. 31, 2019), https://www.bloomberg.com/opinion/articles/2019-01-31/facebook-users-should-be-free-to-sell-their-personal-data.
[129] *See* Artificial Intelligence Forum of New Zealand, Artificial Intelligence: Shaping a Future New Zealand 61 (2018).

especially when the most sensitive data, either commercial or personal, can yield the greatest benefits.[130]

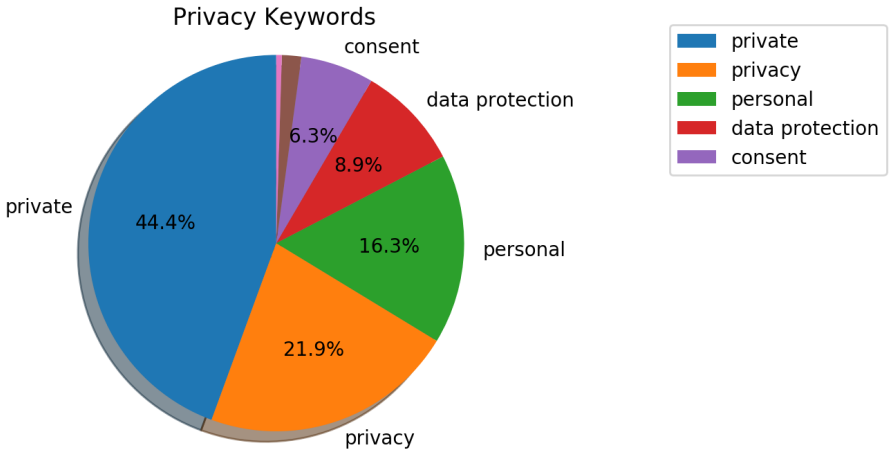

Figure 4: *Privacy Keyword Representation*

Nonetheless, data has long been a commodity for many companies to utilize for personalized advertising and other AI applications. However, the amount of information being collected quickly became a concern of the public following the Snowden revelations on how much data the United States government was collecting and after the previous data regulations outlined by Privacy Shield proved to be ineffective. Shortly thereafter, the General Data Protection Regulation (GDPR) was implemented in Europe, which protected the rights of European citizens to understand what information was being gathered, to access this information, to correct this information, and to erase the personal information which has been gathered about them.[131] AI developers operating in Europe will therefore need to consider this existing policy when creating their AI systems.

---

[130] *See* DEPARTMENT FOR DIGITAL, CULTURE, MEDIA & SPORT & DEPARTMENT FOR BUSINESS, ENERGY & INDUSTRIAL STRATEGY, GROWING THE ARTIFICIAL INTELLIGENCE INDUSTRY IN THE UK 44 (2017).

[131] In the EU, for example, the General Data Protection Regulation (GDPR) provides European citizens with access to and control over what data is collected and how it is used. Commission Regulation 2016/679, 2016 O.J. (L119) (discussing the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC). The EU has also Charter of Fundamental Rights and Freedoms with Article 7 providing that "everyone has the right to respect for his or her private and family life, home and communications," and Article 8 ensuring that "everyone has the right to the protection of personal data concerning him or her." Charter of Fundamental Rights of the European Union, art. 7–8 2010 O.J. (C 83/389). Complete Guide to GDPR Compliance, https://gdpr.eu (last visited Jan. 6, 2021).

Some countries advocate for the availability of data over the protection of privacy. Finland plans to release a legislative framework to ensure the availability of data to business operations, as opposed to focusing on data protection.[132] China and the United States mention protecting privacy in passing in their AI policy documents, but both have limited legal protection of privacy. China does not require explicit consent in most scenarios to collect and process data on consumers.[133] Russia's AI policy does not mention privacy rights at all. The question therefore remains for nations with existing privacy protection frameworks on how to best translate these regulations into AI. One of the nations which possesses the strictest privacy protection laws is South Korea, with their Personal Information Protection Act (PIPA), a predecessor to the GDPR.[134] South Korea's AI framework presents a tripartite policy on data privacy.[135] The nation which discusses the need for consumer privacy the most out of all the nations surveyed is Canada, which has implemented their own privacy protection laws, a GDPR equivalent known as the Personal Information Protection and Electronic Documents Act (PIPEDA). Canada's policy recognizes privacy as a fundamental human right,[136] taking the time to clarify how AI must be utilized to comply with existing privacy principles:

- *Organizations should be required to disclose the use of AI tools + AI-powered services should be opt-in*

---

[132] MINISTRY OF ECONOMIC AFFAIRS AND EMPLOYMENT OF FINLAND, FINLAND'S AGE OF ARTIFICIAL INTELLIGENCE: TURNING FINLAND INTO A LEADING COUNTRY IN THE APPLICATION OF ARTIFICIAL INTELLIGENCE 44(A) (2017).

[133] JACK M. BALKIN, FREE SPEECH IN THE ALGORITHMIC SOCIETY: BIG DATA, PRIVATE GOVERNANCE, AND NEW SCHOOL SPEECH REGULATION (2018).

[134] The European Commission announced in November 2020 that it had presented a new regulation on data governance. In particular, the Commission outlined that the regulation will facilitate data sharing across the EU and between sectors to create wealth for society. *See EU: Commission Presents New Data Governance Regulation*, OneTrust (Nov. 25, 2020), https://www.dataguidance.com/news/eu-commission-presents-new-data-governance-regulation.

[135] The first level would be for general data, which includes no private information, and this data would essentially be available for open-source usage. The second level contains personal information but anonymizes it so that individuals cannot be identified. This data will be used to launch 'free data zones' which allows data corporations to perform data synthesis. The third level contains private data, which is implemented into K-MyData, "a government program that allows businesses to share the personal information of their clients with other businesses, subject to clients' consent." *See* GOVERNMENT OF THE REPUBLIC OF KOREA, MID- TO LONG-TERM MASTER PLAN IN PREPARATION FOR THE INTELLIGENT INFORMATION SOCIETY: MANAGING THE FOURTH INDUSTRIAL REVOLUTION 34 (2017).

[136] *See* OFFICE OF THE PRIVACY COMMISSIONER OF CANADA, CONSULTATION ON THE OPC'S PROPOSALS FOR ENSURING APPROPRIATE REGULATION OF ARTIFICIAL INTELLIGENCE 4 (2020).

- *Create a public complaint and reporting structure for the use of non-evidence-based algorithms, with an option for a formal response from the subject of the complaint*
- *Require the anonymization of individual data when shared publicly to protect privacy*
- *Develop a robust appeal process for those who feel they have been wrongly assessed*[137]

When discussing privacy rights in the context of AI, Canada not only mentions the need to comply to current privacy regulations but discusses the need to modernize and strengthen these regulations to protect users.[138] Finally, the policy even discusses several grey areas regarding consumer privacy, such as when meaningful consent is not practicable.[139]

private	privacy
personal	confidential
control over data	consent
data protection	data governance
protected data	protected information

Table 4, *Privacy Keywords*, September 2020

To summarize, the views upon privacy in relation to AI vary greatly from country to country. Although the amount of time discussing the concept of privacy is generally equally distributed among surveyed nations, there exist many notable outliers such as Canada which drafted a document specifically meant to address privacy concerns within AI, and Russia, which does not mention privacy regulations within their AI strategy. This dimension is likely the most divergent when it comes to its implementation and adoption by different members of the international community, varying greatly upon each nation's existing privacy laws, making global norm building progress difficult.

2.2.5 Fairness

Fairness is a relatively new topic of interest in the AI community. In general, fairness is the state, condition, or quality of being fair, or free from bias

[137] *See* SARAH VILLENEUVE, GAGA BOSKOVIC & BRENT BARRON, REBOOTING REGULATION: EXPLORING THE FUTURE OF AI POLICY IN CANADA 6 (2019).
[138] *See Id*. at 8.
[139] *See* OFFICE OF THE PRIVACY COMMISSIONER OF CANADA, *supra* note 136, at 8.

or injustice,[140] but within machine learning it takes on special meaning. In the AI community, a given algorithm is said to be 'fair' if its results are independent of certain sensitive variables such as gender, ethnicity, sexual orientation, or disability status.[141] Nonetheless, the discussion on fairness is not without complication, as training AI algorithms to be fair relies heavily on non-biased datasets, which may be difficult to identify and qualify as unbiased. Maintaining fairness requires continued input from academia and industry to identify and address current bias issues trending within datasets and algorithms. There have been attempts to create tools that identify bias within algorithms, such as Facebook's Fairness Flow.[142] However, since the source code for Fairness Flow has not been released, it is impossible to determine whether this tool truly corrects bias. Statistical means of detecting discrimination in algorithms also exist, but these have yet to be deeply researched and integrated into a broader AI strategy.

The two keywords that appeared most frequently in the documents surveyed were 'ethical' and 'ethics.' These terms present an idea that is much vaguer than some other keywords, as fairness is assumed to be ethical, but what is ethical may not always be referring to fairness specifically. As it stands, many entities use 'ethical' as a catch-all term to justify their policy decisions, which may not always be supporting fairness. While it is important to take an ethical approach to AI, our definition of fairness is more accurately represented by other keywords such as bias, diversity, inclusion, and discrimination.

---

[140] 'Fairness' Definition, https://www.merriam-webster.com/dictionary/fairness (last visited Jan. 6, 2021).

[141] *Fairness, Accountability, and Transparency in Machine Learning*, FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY IN MACHINE LEARNING, https://www.fatml.org/ (last visited Jan. 19, 2021).

[142] Dave Gershgorn, *Facebook Says It Has a Tool to Detect Bias in Its Artificial Intelligence*, QUARTZ (May 3, 2018), https://qz.com/1268520/facebook-says-it-has-a-tool-to-detect-bias-in-its-artificial-intelligence/; Angie Raymond et al., *Building a Better HAL 9000: Algorithms, the Market, and the Need to Prevent the Engraining of Bias*, 15 NORTHWESTERN J. TECH. & TECH. PROP. 215, 216 (2018).

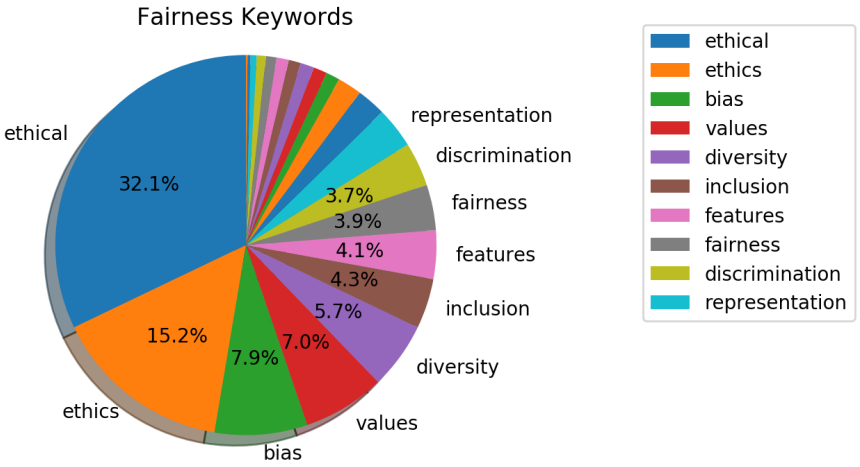

Figure 5: *Fairness* Keyword Representation

While there are many policy manifestos that use the term 'ethical,' most of them skirt the issue of fairness and discrimination. Most policies, such as Sweden's, include a paragraph discussing how biased datasets can lead to discrimination and a loss of trust,[143] but refrain for delving further into the issue. Although a standalone category, fairness is often related to two additional dimensions—transparency and privacy. Several documents discuss the need for algorithmic transparency to ensure fairness, with the United States citing that the lack of transparency and biased data can lead to financial damages and consequences for democratic functions.[144] Nonetheless, a certain level of opacity is inevitable within machine learning algorithms, which can provide high levels of accuracy but do not explain why any classification was made. The need for fairness is also discussed through the lens of privacy, since many researchers have already noted that "existing stereotypes and prejudices against certain groups will be reproduced in the data used by such technologies, leading to unfair and discriminatory decisions."[145] There are some notable exceptions to this trend,

[143] *See* MINISTRY OF ENTERPRISE AND INNOVATION, NATIONAL APPROACH TO ARTIFICIAL INTELLIGENCE 4 (2019).
[144] SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE OF THE NATIONAL SCIENCE & TECHNOLOGY COUNCIL, THE NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN: 2019 UPDATE 19 (2019).
[145] Joanna J. Bryson, Mihailis E. Diamantis & Thomas D. Grant, *Of, for, and by the People: The Legal Lacuna of Synthetic Persons*, 25 A.I. AND L. 273, 273–291 (2017), https://doi.org/10.1007/s10506-017-9214-9

however, such as Mexico's AI White Paper and the Villani Report,[146] which place a special emphasis on how AI affects gender equality in the workplace.[147]

In addition to the overlap with transparency and privacy, fairness is also deeply connected with public benefit. Fairness within AI impacts the ability of all members of the public to reap in the benefits provided by AI, but this is especially important to marginalized populations. It is widely documented that black faces are underrepresented in facial recognition software,[148] and when a vending machine which operates on facial recognition is unable to recognize a black customer, the technology is rendered to be inaccessible. Many countries reference the importance of access to data, but a lot of discussions on access are framed from a business perspective rather than a social one. There are some countries, such as Italy, which emphasize the need to utilize AI to reduce inequalities in the healthcare and education sector.[149] The Villani Report takes this a step forward, specifically acknowledging that algorithms reinforce bias against women and black communities, which limit their accessibility to employment, housing, and access to goods and services.[150]

Overall, the most comprehensive discussions on fairness and bias through the lens of AI can be found in the Villani Report and 'Shaping a Future New Zealand.'[151] New Zealand brings up a specific case study from the United States: an AI system was being used by judges to set sentencing in a way to reduce the risk of repeat offenses, but because this algorithm relied on historical data, it developed a bias against black defendants, leading to longer prison sentences.[152] New Zealand's policy proposes to curb this bias by diversifying the pool of AI developers and the datasets upon which AI are trained.[153] France's policy mentions similar situations where women are offered lower paid jobs by Google and proposes to rectify these inequalities through legal frameworks and auditing systems which can identify and quantify bias.[154]

---

[146] CÉDRIC VILLANI, FOR A MEANINGFUL ARTIFICIAL INTELLIGENCE: TOWARDS A FRENCH AND EUROPEAN STRATEGY 140 (2018).
[147] TOWARDS AN AI STRATEGY IN MEXICO: Harnessing the AI Revolution at 28
[148] Spencer Buell, *MIT Researcher: Artificial Intelligence Has a Race Problem, and We Need to Fix It*, BOSTON MAG. (Feb. 23, 2018), https://www.bostonmagazine.com/news/2018/02/23/artificial-intelligence-race-dark-skin-bias/.
[149] *See* AGENCY FOR DIGITAL ITALY, 2019 DRAFT STRATEGY FOR ARTIFICIAL INTELLIGENCE 9 (Artificial Intelligence Task Force ed., 2018).
[150] Villani, *supra* note 146, at 116.
[151] *Id*.
[152] Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. and It's Biased Against Blacks.*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
[153] Artificial Intelligence Forum of New Zealand, *supra* note 129, at 64.
[154] Villani, *supra* note 146, at 116.

Nonetheless, there are some unique aspects of fairness which are not seen in France's and New Zealand's policies. Germany's AI policy includes a subsection specifically relating to fairness within culture and media. The manifesto states that even in the AI age, the "freedoms of a democratic society will still primarily be measured in terms of cultural and media diversity and the independence of the media."[155] While AI will not replace human creativity, it will play a vital role in providing an environment for media consumption. In order to protect this environment so that free speech can flourish, AI must abide by the principles of transparency and non-discrimination.[156]

Discussions on equity are notable absent from any of the documents we have surveyed. While the keyword is occasionally used in the text, none of the documents attempt to discuss the difference between equality and equity. There are statistical tools which can be used to identify bias within algorithms, but the idea of fairness is not rooted merely in the absence of bias. Bias can be used within algorithms to create equitable outcomes. The MIT-D lab presents a hypothetical situation in which an algorithm was more likely to disqualify women from receiving a business loan over men, regardless of creditworthiness. A bias can be introduced to the algorithm to rate women's creditworthiness higher than men to alleviate the inequality faced by women entrepreneurs.[157] Thus, the introduction of a bias in this case could be said to be promoting fairness.

Fairness is an incredibly important aspect of artificial intelligence but is vastly underrepresented in international AI policy. Most policies converge on the idea that fairness is a crucial consideration to machine learning algorithms, yet they rarely expand on the topic to discuss how this fairness will be measured and ensured. Discussions of fairness are shallow and limited, and the topic is overrepresented in our dataset due to the abundance of the keyword 'ethical', which does not always refer to algorithmic fairness and protection against discrimination.

Table 5: *Fairness Keywords*

fairness	ethics
ethical	unethical
discriminate	discriminatory

[155] FEDERAL MINISTRY OF EDUCATION AND RESEARCH, FEDERAL MINISTRY FOR ECONOMIC AFFAIRS AND ENERGY & FEDERAL MINISTRY OF LABOR AND SOCIAL AFFAIRS, ARTIFICIAL INTELLIGENCE STRATEGY: AI MADE IN GERMANY 44 (2018).
[156] *Id*.
[157] YAZWEED AWWAD, RICHARD FLETCHER, DANIEL FREY, AMIT GANDHI, MARYAM NAJAFIAN & MIKE TEODORESCU, EXPLORING FAIRNESS IN MACHINE LEARNING FOR INTERNATIONAL DEVELOPMENT 2 (2020).

discrimination	bias
debiasing	inclusion
diversity	equality
equal	equitable
democratic	representation
objectivity	inclusiveness
non-discrimination	values
features	harm
mitigation	data quality

2.2.6 Human-Centered Design

Human-Centered Design is an approach to problem solving, commonly used in design and management frameworks that develops solutions to problems by involving the human perspective in all steps of the problem-solving process. Humans have become the measuring stick for AI performance, with many documents noting that while AI can outperform humans at a specific task, the intelligence associated with humans cannot be reproduced in AI at this time.[158] Instead, the best way to utilize AI would be to use them as prosthetics to enhance human cognitive ability. Big data and algorithms alone cannot accurately gauge reality without specific guidance, and human-centered design is crucial to avoid biased algorithms and a reduction of effectiveness.[159] Ultimately, human-centered design is at the core of the creation of AI, as humans are creating AI to serve their needs, rather than the other way around. When referring to human-centered design, we must additionally keep in mind societal norms which shape our ideas of how AI should function in society, and different nations will likely have different mindsets of how AI should interact with humanity.

To address this broad spectrum of concerns, we used the keyword 'human' to track the amount of time dedicated to discussing human-centered design, and relatedly human rights. This was by far the most widely used keyword within all chosen keywords across different dimensions, numbering over 500 uses throughout all our chosen documents. While this may initially seem to be too general, every time an AI interacts with humans, human-centered design is a key consideration within this interaction.

---

[158] *Id*. at 19.

[159] Jim Guszcza, *Smarter Together: Why Artificial Intelligence Needs Human-centered Design*, 22 DELOITTE REV. 36 (2018).

Another keyword that appeared was 'augmented,' referring to the term 'augmented intelligence.' This terminology refers to augmenting artificial intelligence with Human-Centered Design components to better adjust to human expectations to achieve the goals set by the creators of the AI. In this sense, the artificial intelligence plays an assistive role to human cognition.[160] Machine learning, on its own, lacks what we would call 'common sense,' and this augmentation is necessary to bridge the gap between humans and machines.

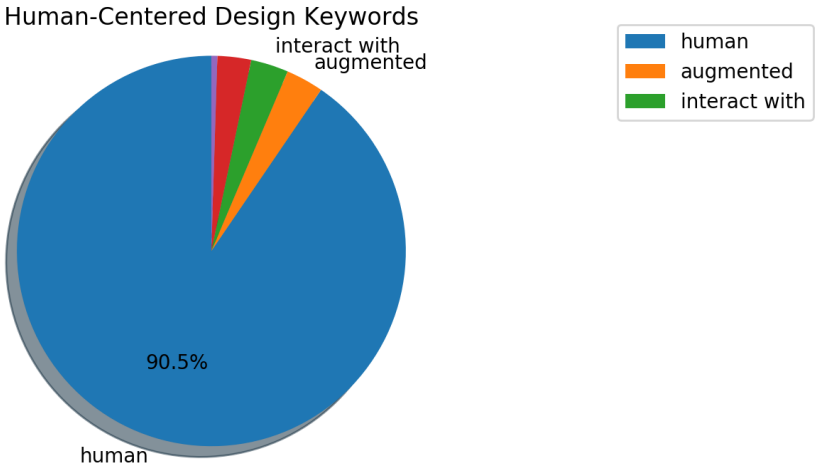

Figure 6: ***Human-Centered Design Keyword Representation***

We have found that human-centered design manifests in policies in several different forms. First, there is the need for AI to be integrated into society in a way that humans will accept. Secondly, there is a need to set up a standard of fundamental human rights that shape AI development and policy. Thirdly, following the integration of AI into society, there is the need for human oversight and control over AI. Finally, we have found that there is the emphasis on adapting AI to resemble human-beings, creating a special subset of human-AI interaction.

The three countries that spend the most time discussing human-centered design in our study were the United States, Austria, and Japan. Japan represents a bit of an outlier, as the majority of the keyword count comes from the term 'human resources,' which refers to the people representing the workforce of an organization rather than the interaction of humans with AI. Nonetheless, Japan still mentions the necessity of social

[160] *Augmented Intelligence*, GARTNER, https://www.gartner.com/en/information-technology/glossary/augmented-intelligence (last visited Jan. 20, 2021).

acceptance of AI and the need for AI to be 'human-friendly.'[161] Italy, another country which focuses on the idea of human-centered design lays out a similar thought upon the inclusion of AI into society: "laws, regulations and good technical and technological practices are not enough (though necessary): we need a narrative and an imaginary built by society in an inclusive way outlining the meanings of AI and the roles we want to assign to it."[162] Not all countries address the need for this narrative directly, but the fundamental concepts behind it can be summarized as following: the need for AI to acknowledge basic human rights, the amount of control humans retain over AI decisions, and the humanization of AI.[163]

The first aspect of the narrative lies in the necessity for AI to observe human rights. Some countries specifically mention fundamental human rights within their AI strategy, such as Austria. Austria's framework principles state that "[t]he use of robotics and AI must guarantee the safety of people and comply with ethical standards, fundamental human rights and European values,"[164] meaning that AI must be created with human rights at the forefront of their development. These human rights include "human freedom and dignity, economic, cultural and social rights and the protection of privacy."[165] Italy also focuses on the protection of human rights, citing that 'human well-being is the highest virtue for a society,' and to necessitate this, certain inalienable human rights are required.[166] It should be noted that this discussion is absent from some policy manifestos, and but this omission cannot be definitively stated as an objection to fundamental human rights or the need for AI to observe them.

Another cornerstone issue commonly addressed in these policies was the amount of influence humans have over AI decision-making processes. This topic shares a significant overlap with the dimension of Accountability, yet the foundation of this issue lies within human-machine interaction. Many AI units have manual override options incorporated in their design to account for human agency. Italy's policy suggests that AI should be primarily used to improve human judgement, rather than to override it completely.[167] Austria's manifesto outright states that "[m]achines cannot and should not assume moral responsibility. Ethical

---

[161] *See* GOVERNMENT OF JAPAN, INTEGRATED INNOVATION STRATEGY 73 (2018).
[162] *See* Agency for Digital Italy, *supra* note 150, at 32.
[163] *Id*.
[164] *See* Austrian Council on Robotics and Artificial Intelligence, *supra* note 161, at 6.
[165] *Id*.
[166] *See* Agency for Digital Italy, *supra* note 150, at 9.
[167] *Id*. at 24.

responsibility for robotics and AI systems must ultimately remain with humans."[168] We have found that many documents come to the same conclusion if they choose to address this topic. Interestingly, China's policy specifies utilizing a hybrid approach to human-machine decision making, such as utilizing human-machine collective driving.[169]

The last aspect of human-centered design addresses the humanization of AI. Although not all countries have a specific section addressing this topic, the policies set forth by the United States and Austria contain a special emphasis places upon creating humanoid robots. The Austrian manifesto claims that "anthropomorphic design features such as faces, eyes and arms are used to transmit non-verbal signals known from interpersonal communication to the robot, thus promoting a "social" connection with the machine."[170] This provision acts as a catalyst to integrate AI into human society and culture, and is further backed by social science literature, which "suggests that people—depending on personal and situational factors—tend to perceive intentionality or other human characteristics in robots and AI systems."[171]

human	people oriented
psychology	interact with
HCI	augmented
operationalized	

Table 6: *Human-Centered Design Keywords*

In our research, we have found that the countries who choose to discuss human-centered design tend to converge on the same topics, though not generally on human rights. The divergence is noted mostly in the omission of the topics themselves, which may be a result of prioritization of different AI principles rather than espousing different ideologies upon human-AI interaction.

2.2.7 Public Benefit

In general, public benefit is a benefit accrued to the public. For example, enhanced mobility of people or goods, environmental protection

[168] Austrian Council on Robotics and Artificial Intelligence, *supra* note 161, at 25.
[169] STATE COUNCIL OF CHINA, NEXT GENERATION ARTIFICIAL INTELLIGENCE DEVELOPMENT PLAN ISSUED BY STATE COUNCIL 12 (2017).
[170] *See* Austrian Council on Robotics and Artificial Intelligence, *supra* note 161, at 36.
[171] *Id*.

or enhancement, congestion mitigation, enhanced trade and economic development, improved air quality or land use, more efficient energy use, enhanced public safety or security, and similar benefits that accrue to the public. In the context of AI, public benefit can take on two different meanings. In the context of government use of AI, public benefit can be AI that is designed to achieve a public benefit- or enhance its occurrence- using AI. In the context of private use, public benefit can mean the use of AI in a manner that – despite private actor – seeks to benefit the public through deployment.

'Public' is the most popularly used keyword in the documents surveyed, but like many terms, not every instance of 'public' refers directly to public benefit, or human rights. Often 'public' can be used to refer to the 'public sector,' which is only tangentially related to public benefit. Several term that was not included in the list of keywords were 'transportation,' 'utilities,' and 'agriculture,' but they frequently came up in discussions on how AI would benefit the public. Nations would frequently group several of these terms together to discuss public benefit, such as Denmark presenting the grouping of 'healthcare, energy and utilities, agriculture, and transport.'[172] Different nations chose to focus on varying aspects of public benefit, but the most common benefits described were healthcare, education, agriculture, energy and utilities, transportation, and environment. Although 'economy' was included as one of the terms in public benefit, many discussions about economy concerned only the business side of AI rather than its role as a public benefit and thus will not be covered in this section.

[172] *See* MINISTRY OF FINANCE & MINISTRY OF INDUSTRY, BUSINESS AND FINANCIAL AFFAIRS, DANISH NATIONAL STRATEGY FOR ARTIFICIAL INTELLIGENCE 63 (2019).

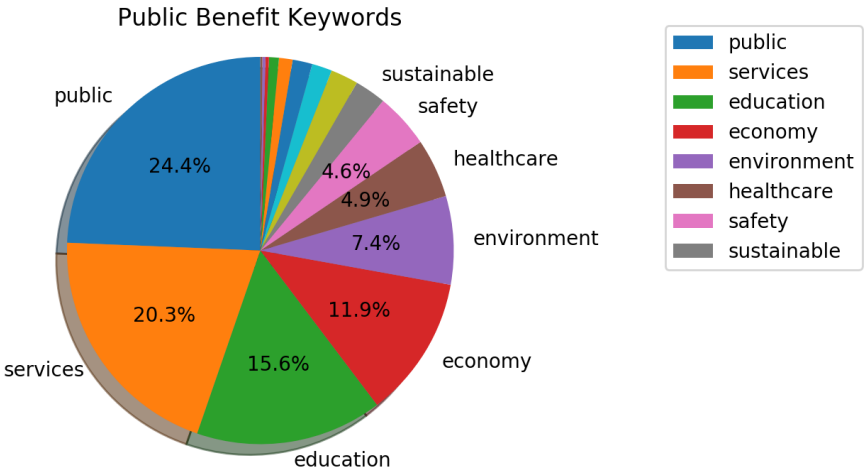

Figure 7: ***Public Benefit Keyword Representation***

Healthcare was one of the most frequently covered examples of the public benefit that AI will bring. It is slightly underrepresented in Figure 7 due to some countries referring to this issue as 'healthcare' or 'health management,' or simply 'health.' Some nations, such as India, noted that there was a lack of access to healthcare in their country, and AI would help many people get access to these services who may not have had access to them before.[173] Furthermore, the induction of AI into the healthcare system will help with inefficiency and save money. New Zealand noted in its policy that "algorithms could save the global health sector up to US$100 billion a year, as a result of AI assisted efficiencies in clinical trials, research and better decision making in the doctor's office."[174] Many nations that are facing an aging demographic noted that AI can be utilized to care for elderly populations with "robotic systems designed for use in in-patient and out-patient care or in people's homes . . . Smart robotic systems can be used for treatment purposes, communication and interaction, moving people and helping them stay mobile, assisting and accompanying them."[175] In addition to these benefits, AI also excels at chronic disease prevention and management by "analyzing clinical data, health behaviors, and genomic data to create personalized risk scores for

[173] *See* NITI AAYOG, NATIONAL STRATEGY FOR ARTIFICIAL INTELLIGENCE #AIFORALL 24 (2018).
[174] *See* Artificial Intelligence Forum of New Zealand, *supra* note 128, at 59.
[175] *See* Federal Ministry of Education and Research, Federal Ministry for Economic Affairs and Energy & Federal Ministry of Labor and Social Affairs, *supra* note 159, at 18–19.

individuals,"[176] promoting the human right to health, as noted in the policies for both India and New Zealand.

AI has also been found to assist in education. While AI cannot replace the teacher directly, it can "greatly assist teachers in efficiently and effectively managing multi-level / multigrade classrooms, by judging learning levels of individual students, and allowing automated development of customized educational content adapted to each child's class and learning level."[177] Italy has invested in developing automated tutoring, personalized learning, and even has AI calculating drop-out risk through predictive indicators.[178] Education has also been discussed thoroughly in advancing AI knowledge among the population to empower citizens on how to best take advantage of the benefits AI has to offer.

As many nations still rely on agriculture as an integral part of their economy, AI is also being used to promote efficiency in this industry, and thus indirectly promote the human right to food.[179] India is investing in robotic technologies to produce agribots, which will weed fields, fertilize plants and harvest crops.[180] Countries that struggle with limited amount of water, such as India and New Zealand, found that AI can promote effective water usage (another human right).[181] Advanced sensors can "keep track of the soil status and weather forecast to prevent under and over irrigation, helping alleviate water usage inefficiency."[182]

The increased efficiency would also benefit the energy and utility sector. Singapore's AI policy mentions that fault detection and maintenance management in critical infrastructure can reduce the risk of system failure.[183] Lithuania's policy notes that AI can be used to create more efficient ways to distribute power, and therefore decrease their reliance on foreign sources of energy while increasing sustainability.[184] The focus on sustainability is also noted in other country policies; Italy,

---

[176] *See* SMART NATION AND DIGITAL GOVERNMENT OFFICE, NATIONAL ARTIFICIAL INTELLIGENCE STRATEGY 30 (2019).
[177] *See* Aayog, *supra* note 175, at 37.
[178] *See* Agency for Digital Italy, *supra* note 150, at 22.
[179] *See The Right to Food*, UN Food & Agricultural Org., http://www.fao.org/right-to-food/en/ (last visited Jan. 6, 2021).
[180] *See* Aayog, *supra* note 175, at 32.
[181] *Human Right to Water*, https://www.un.org/waterforlifedecade/human_right_to_water.shtml (last visited Jan. 6, 2021).
[182] *See* Aayog, *supra* note 175, at 32.
[183] *See* Artificial Intelligence Forum of New Zealand, *supra* note 128, at 47.
[184] LITHUANIAN ARTIFICIAL INTELLIGENCE STRATEGY: A VISION OF THE FUTURE 10, 15 (2019).

Japan, and France have a special focus dedicated to environmental sustainability.

AI monitoring systems can not only promote sustainability, but also track any potential damage to the environment and deploy appropriate solutions. Japan, on the other hand, pledges to use AI to "practice energy/climate change diplomacy... dealing with climate change and improving energy security by means of assisting other countries' initiatives to achieve SDGs mainly with the use of low-carbon type infrastructure technologies including renewable energies and hydrogen."[185] France takes the opportunity to use "AI in the field of ecology: AI can help us understand the dynamics and the evolution of whole ecosystems by focusing on their biological complexity; it will allow us to manage our resources more efficiently (particularly in terms of energy), preserve our environment and encourage biodiversity."[186] Increasing efficiency to reduce greenhouse gases is a crucial idea to yet another public sector—that of transportation. As well as promoting the reduction of greenhouse gases, India's AI policy proposes using semi-autonomous vehicles to reduce congestion and fatalities on the road.[187] South Korea's AI policy boasts several innovations that would assist with transportation including preventative maintenance on vehicles, parking spot finders, and real time data for traffic reduction.[188]

AI can also be used to increase security. This has been briefly discussed earlier in the 'Security' dimension, but it must be noted that China focuses on security as a public benefit derived from AI. Singapore also lists security as one of the benefits which come from AI, although their policy has a specific focus: that of border control. Singapore aims "to deploy AI to achieve 100% automated immigration clearance for all travelers, including first-time social visitors. Singaporeans and departing visitors will experience "BreezeThrough" immigration clearance, without the need to present their passports."[189] This will be achieved through Singapore Arrival Cards submitted by travelers and passenger information from various airlines.[190]

---

[185] *See* Government of Japan, *supra* note 164, at 91.
[186] *See* Villani, *supra* note 147, at 102.
[187] *See* Aayog, *supra* note 175, at 43.
[188] *See* GOVERNMENT OF THE REPUBLIC OF KOREA, *supra* note 135, at 13-15.
[189] *See* NATIONAL ARTIFICIAL INTELLIGENCE STRATEGY: SINGAPORE 35, https://www.smartnation.gov.sg/why-Smart-Nation/NationalAIStrategy.
[190] *See id*.

public	safety
welfare	economy
economics	healthcare
education	services
improvement	environment
stakeholder	peace
sustainable	sustainability
climate	mitigating risk
benefits of	general interest

Table 7: *Public Benefit Keywords*

Finally, all these developments play into the realization of smart cities. Smart cities are "being developed are trying to solve challenges such as low visibility on usage of utilities such as electricity, water, and waste management."[191] They also meet the demands of a rapidly growing urban population, resulting in a better quality of life. Several nations discuss the creation of smart cities in their AI Strategies, including India, South Korea, and Singapore, although many nations are not quite ready to commit to such a usage of AI. Overall, there is a lot of convergence on utilizing AI for public benefit, although many nations have unique ideas on how to do so complicating the path for AI norm development.

2.3 Summary

Our research showed global convergence of patterns on multiple AI themes, in particular for the public benefit. This convergence provides a strong impetus for global collaboration and norm creation in this domain. The most popular public benefits brought about by AI according to the surveyed documents were in the healthcare sector, education, and transportation. In this case, many countries focused on their own nation's unique context.

[191] *See* Villani, *supra* note 147, at 39.

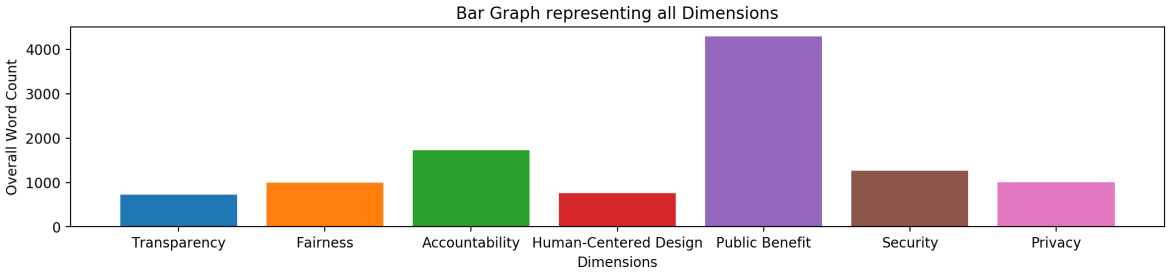

Figure 8: ***Word Count for Each AI Dimension***

Similar patterns were found in other categories as is shown in Figure 9. The topic of security was present in almost every policy we examined, but the attitudes of utilizing AI to promote security were varied. China, Singapore, and South Korea wanted to use AI to promote security, but other nations such as France and Canada highlight how security using AI can lead to mass surveillance and the deterioration of personal autonomy. In this case, AI policy reflected each country's cultural attitudes towards the concept of collective security versus individual privacy. In general, however, we have found that countries spent most of their paper discussing the public benefit of AI and how to ensure accountability. Security, privacy, fairness, transparency, and human-centered designs were themes that were touched upon, but they were overshadowed by the other two categories by a wide margin.

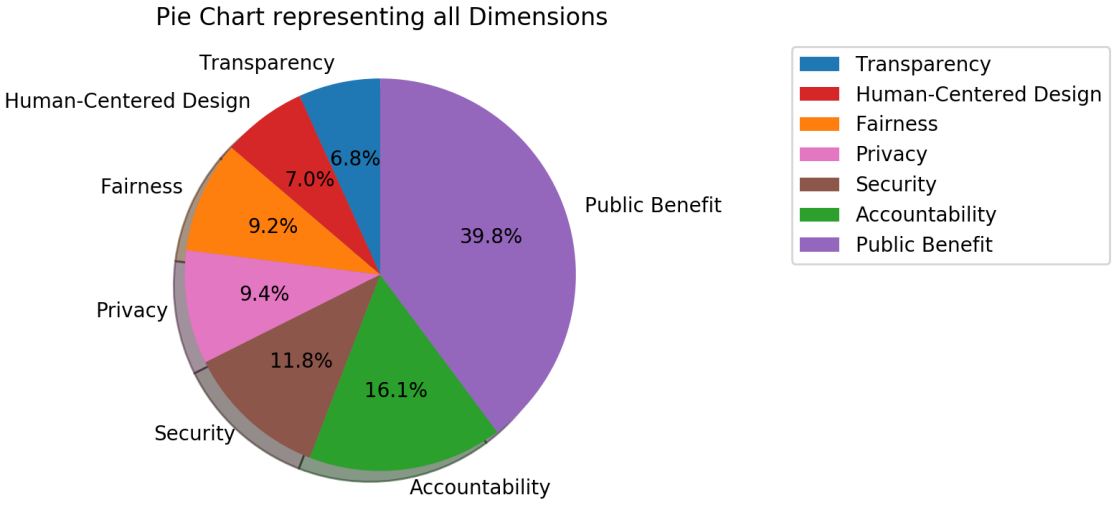

Figure 9: ***Word Count Per Dimension***

In addition to our findings, we have confirmed the convergence by examining the standard deviation of our findings as is summarized in Figure 10. While public benefit had the highest standard deviation out of

all the dimensions, this standard deviation did not exceed ten percent of the overall word count. The smaller six categories had even less variation in terms of percentage, leading to a consistent global standard in word count dedicated to discussing these issues.

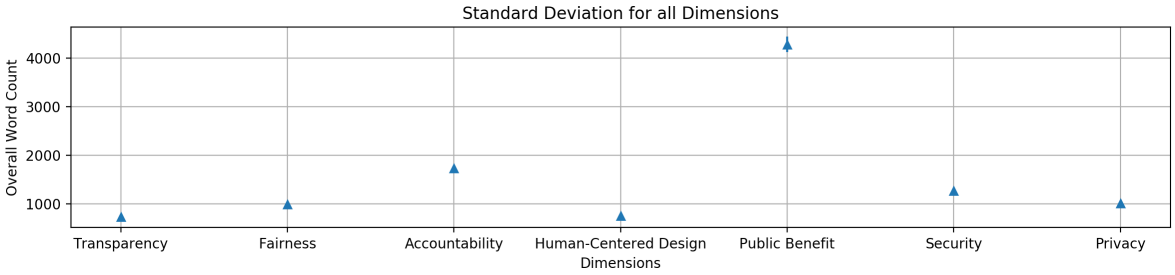

Figure 10: ***Graph showing Standard Deviation for all Dimensions***

3. IMPLICATIONS FOR PRACTITIONERS AND POLICYMAKERS

The findings from Part 2 should merely be used as a starting point for broader discussions within the area of AI norm development pertaining to human rights protections. This section draws upon the above survey and corresponding conclusion make specific recommendations as to next steps. First, this Part explores the limitations of the study, and then continues by highlighting the areas of convergence highlighted above. The Part concludes with a discussion criticisms that arise within the AI strategies themselves, and concludes by making suggestions for future work and considerations.

3.1 Study Limitations

This Article did not explore a variety of categories related to AI given space and conceptual constraints. Specific related technologies such as blockchain, Internet of Things, 5G, and quantum computing were not examined in detail. These microcosms contain their own parallel discussions on cybersecurity, privacy, and other dimensions within themselves, and thus were only considered in the context that they added depth and nuance to each dimension surveyed. Attempting to discuss each specific technology, technique, or application would likely merit a paper of its own. For example, we did not discuss the topic of AI being used in defense, though the issue does come up in both France and Russia's AI

Strategies.[192] The United States, moreover, has spent billions on unclassified and classified AI R&D for the military, and in 2018 the Joint Artificial Intelligence Center was established to oversee U.S. defense agency AI research.[193]

Industry collaboration and public-private AI partnership is another sector that we did not consider. Although it belongs to the category of public benefit, this is a subject that deserves special attention given the extent to which the public sector can influence the private sector, and vice versa, in different nations. Idiosyncrasies can arise due to the varied norms and expected levels of government oversight in various nations. In the EU, many countries have revealed their plans to share government research with private firms in support of startups. The United Kingdom is catalyzing "over £300 million in private sector [AI] investment."[194] Japan has taken it one step further, integrating the private sector into the development of their AI strategy.[195] France has described their plans to create four or five research facilities, with Germany joining in to help fund a Franco-Germany research and development collaboration.[196] These distributed, multi-stakeholder approaches to coordinate AI governance could be considered polycentric, as is noted below, though that does not necessarily make them more robust or successful than other governance models.

As is evident, we hope that this brief survey of the current state of AI strategies and the core dimensions that many of them touch on is among the first, but certainly not the last, word on the topic. Follow-up can and should discuss the interrelationships between these arenas, how they are being operationalized in domestic policy, and whether or not the entire enterprise of crafting separate strategies for AI, cybersecurity, data governance, etc. is in fact creating artificial separations that are making it more challenging for nations to tackle such deeply connected, multifaceted domains, as is discussed further below.

3.2 Taking Stock of AI Norm Development

---

[192] *AI Strategies*, https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd (last visited Jan. 7, 2021).
[193] *Id*.
[194] *Id*.
[195] *Id*. ("The 11-member council which created the Public-Private Dialogue towards Investment for the Future had "representatives from academia, industry, and government, including the President of Japan's Society for the Promotion of Science, the President of the University of Tokyo, and the Chairman of Toyota.").
[196] *Id*.

As is detailed in Section II, there is an extensive amount of movement in the field of AI strategy and governance. Globally, governments, private sector organizations, and civil society groups continue to highlight the economic potential of AI while also noting the importance of developing standards around how to use and deploy this revolutionary technology. Many regulators are considering how to encourage innovative uses of AI while also preserving and protecting human rights and societal values. One benefit of examining the numerous national AI strategies and initiatives listed above is to identify areas of convergence and potential for norm-building. Even with the vast array of stakeholders involved and the complexities of legal, ethical, and technical questions around AI, there seems to be considerable agreement around many of the principles examined in Part 2. The dimensions surveyed are not only convergent among national strategies, particularly in the public benefit context, but they are also emerging in global and regional intergovernmental initiatives to develop and promote AI.

Perhaps the most well-recognized and wide-reaching of these intergovernmental initiatives is the OECD's Recommendation of the Council on Artificial Intelligence, which highlights five "values-based principles" to promote the "responsible stewardship of trustworthy AI."[197] These principles are: "inclusive growth, sustainable development and well-being; human-centered values and fairness; transparency and explainability; robustness, security and safety; and accountability."[198] The principles are coupled with recommendations on how to invest in and encourage AI development. The OECD "Principles for Responsible Stewardship of Trustworthy AI" have forty signatories and were endorsed and adopted by the G20 (which notably includes China and Russia). The G20 Statement highlighted the potential of the principles to help with "maximizing and sharing the benefits from AI, while minimizing the risks and concerns."[199] These principles currently stand as the most well-developed international document of principles applicable to AI development.

The focus on building trustworthy AI is also at the center of regional efforts to develop principles around AI deployment. The Nordic Council of Ministers released "AI in the Nordic-Baltic Region" in May 2018, which detailed the potential of AI for the region and announced collaboration to enhance access to data, develop ethical guidelines and values, and promote those values within

[197] *Recommendation on the Council of Artificial Intelligence*, OECD (May 21, 2019), https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.
[198] *Id*.
[199] *G20 Ministerial Statement on Trade and Digital Economy*, G20, at 4 (June 2019), https://www.mofa.go.jp/files/000486596.pdf.

the EU.[200] The European Commission created a High-Level Expert Group on AI (AI HLEG), and they released draft "Ethics Guidelines for Trustworthy AI" in 2018.[201] The Guidelines encouraged organizations to consider the following principles when developing and deploying AI: "human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental well-being; and accountability."[202] Using these principles, the AI HLEG published "The Assessment List for Trustworthy Artificial Intelligence (ALTAI)" in 2020, hoping to provide organizations with a self-assessment tool to promote AI in a way that aligns with economic prosperity and innovation while supporting and upholding human rights and other fundamental values.[203]

These are only a few prominent examples of how initiatives continue to focus on prioritizing economic development of AI while ensuring societal and social well-being. The UN has also emphasized the importance of centering human rights and values with AI to promote sustainable development. Most recently, the United Nations Educational, Scientific, and Cultural Organization (UNESCO) appointed a 24-member expert group to "draft internationally applicable recommendations on ethical issues raised by the development and use of AI."[204] The Expert Group met for the first time in April and is expected to develop recommendations throughout the next few months and submit them to UNESCO's Member States at the next general conference.[205] This intergovernmental effort is still in early stages, but this could prove promising for laying the groundwork for AI norm development.

Another intergovernmental effort in early stages is the Global Partnership on AI (GPAI), which is a primarily G7 effort discussed in Part 1 to "support the responsible and human-centric development and use of AI in a manner consistent with human rights, fundamental freedoms, and our shared democratic values, as

---

[200] *AI in the Nordic-Baltic Region*, NORDIC COUNCIL OF MINISTERS (May 14, 2018), https://www.regeringen.se/49a602/globalassets/regeringen/dokument/naringsdepartementet/20180514_nmr_deklaration-slutlig-webb.pdf.

[201] ETHICS GUIDELINES FOR TRUSTWORTHY AI, EU (Apr. 8, 2019), https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

[202] *See Shaping Europe's Digital Future: Artificial Intelligence*, EUR. COMM'N https://ec.europa.eu/digital-single-market/en/artificial-intelligence (last visited Jan. 7, 2021).

[203] *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment*, EUR. COMM'N (July 17, 2020), https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment.

[204] *UNESCO Appoints International Expert Group to Draft Global Recommendation on the Ethics of AI*, UNESCO (Mar. 3, 2020), https://en.unesco.org/news/unesco-appoints-international-expert-group-draft-global-recommendation-ethics-ai.

[205] *Id*.

elaborated in the OECD Recommendation on AI."[206] GPAI members include Australia, Canada, France, the European Union, Germany, India, Italy, Japan, Mexico, New Zealand, the Republic of Korea, Singapore, Slovenia, the United Kingdom, and the United States.[207] GPAI will collaborate with various stakeholders to explore the themes of responsible AI, data governance, the future of work, and innovation and commercialization.[208] Finally, the EU released potentially groundbreaking guidance on new data governance regulations in November 2020, which as of this writing remain in draft form.[209]

These efforts repeatedly emphasize that society stands to benefit greatly from the enormous potential of AI, but there is also a strong emphasis on developing trust in order to promote its adoption. Perhaps this is the rationale behind why public benefit, accountability, and security are so clearly featured in both national and international AI strategies and principles.

3.3 Criticisms

After a full review and the examination of convergence in some key areas of AI strategies and policy, the authors contend that while some convergence is occurring potentially highlighting opportunities to crystallize State practice and aid norm development, it would be unfair to not also discuss concerns that are beginning to emerge.

First, one has to ask – is AI – or any technological tool for that matter, capable of being regulated or even honed by the creation of a national strategy? Moreover, AI refers to a suite of technologies with myriad applications, but left unexamined is what exactly makes AI distinct from other technological innovations? Further, one wonders if a strategy is in fact appropriate, does any strategy succeed without in-depth consideration for how it may be operationalized? There are a litany of tools, strategies, and other efforts of best practice and community building for self-governance, yet the impacts of these efforts – such as national cybersecurity strategies – remains unclear.[210]

Second, the suite of technologies comprising AI and ML does not respect jurisdictional bounds. Thus, there is a legitimate question about the value of such national-level initiatives to address global challenges. There is potential strength

---

[206] *Joint Statement from Founding Members of the Global Partnership on Artificial Intelligence*, U.S. DEP'T. OF ST. (June 15, 2020), available at https://www.state.gov/joint-statement-from-founding-members-of-the-global-partnership-on-artificial-intelligence/.
[207] *Id*.
[208] *Id*.
[209] *See EU: Commission Presents New Data Governance Regulation*, ONETRUST (Nov. 25, 2020), https://www.dataguidance.com/news/eu-commission-presents-new-data-governance-regulation.
[210] *See, e.g.*, Shackelford & Kastelic, *supra* note 17.

with different local, federal, regional, and private-sector stakeholders taking their individual approaches to harnessing the power of AI to promote human rights and sustainable development, but only if such efforts are coordinated.[211] Without such interaction, communication, and coordination there is a higher likelihood of gridlock and a chaotic fracturing of governance, as is made evident by the literature on polycentric governance.

Putting all of that aside, for the sake of argument, there are also fundamental issues with some of the strategies themselves. AI principles and values, if too broadly defined, can become meaningless. Part 2 outlined the prominent definitions of the principles featured in national strategies, but many of the terms have unclear or ambiguous definitions. For example, what does a strategy mean by "fairness," and who gets to define, what is fair?

Moreover, AI strategies are too narrowly focused on one technology, which is a similar issue with cybersecurity strategies; data governance strategies may be more appropriate to get to the heart of the ethical, legal, and social issues involved. Putting up silos around specific technologies is unlikely to reap the expected benefits, and could cause both confusion and lead to reactive policymaking given that policymakers will constantly be chasing the latest technical trend (blockchain, quantum, etc.) rather than focusing on the bigger picture on how all of these technologies are impacting national security, human rights, and sustainable development.

Finally, as has been noted many AI deployments are built on large aggregate data sets, are ubiquitous, focus on averages, are opaque, and often too complex for any regulator to easily monitor. And while AI strategies might state lofty goals of protecting individuals from ever encroaching AI deployments, the simple fact is the technology business model is built on well-understood aspects of human decision-making regarding incentives, rewards for clicks, and a need for interactive growth.[212] Human rights, personal freedoms, and autonomy are not part for the typical commercial model, yet this aspect of AI governance – what some may argue is an AI market failure[213] – is often missed in national AI strategies.

---

[211] *See, e.g.*, Michael D. McGinnis, Elizabeth B. Baldwin, & Andreas Thiel, *When is Polycentric Governance Sustainable? Using Institutional Theory to Identify Endogenous Drivers of Dysfunctional Dynamics* (2020), https://ostromworkshop.indiana.edu/pdf/seriespapers/2020fall-colloq/mcginnis.pdf.

[212] *See e.g.*, Chavie Lieber, *Tech Companies Use "Persuasive Design" to Get us Hooked. Psychologists say it's Unethical*, VOX (Aug. 8, 2018), https://www.vox.com/2018/8/8/17664580/persuasive-technology-psychology.

[213] *See* William Magnuson, *Artificial Intelligence on Wall Street Will Be Great, Until It Isn't*, GOVERNING (Nov. 4, 2019), https://www.governing.com/news/headlines/Artificial-Intelligence-on-Wall-Street-Will-Be-Great-Until-It-Isnt.html.

3.4 Next Steps

We view this analysis as among the first, and certainly not the last word in analyzing the convergence of AI strategies as a useful data source for identifying arenas for norm development particularly in the arena of human rights, as is discussed above. Among other promising candidates for future research, we suggest comparing and contrasting national-level AI strategies with those at the city and state level, particularly in federated democracies such as the United States and Australia. This type of analysis would prove vital in better understanding to what extent multilevel stakeholder engagement is being utilized to build support for a common vision of AI governance in a given society. Such open lines of communication, along with the coordination and interaction that they engender, is vital for trust building and successful polycentric partnerships, as Elinor Ostrom demonstrated.[214]

Relatedly, as was discussed in Section 3.1, it is important to relate public and private-sector AI governance efforts. It would be interesting to see how much overlap there is between the AI-related governance efforts of firms in a particular nation with that country's AI strategy. These data would be another useful proxy for the overall level of multi-stakeholder engagement and related opportunities to engender nested governance structures.

We also recommend a series of case studies in various facets of AI governance, including efforts to control the spread of so-called deep fakes, to apply these findings to various real-world problems. Part and parcel of this analysis will include a deep dive into prevailing global cultural and ethical considerations as they pertain to AI governance strategies, including the role of deontological, teleological, virtue ethics, Confucian ethics, and other global traditions.

There are overarching questions to consider, including whether it makes sense to maintain separate national AI and cybersecurity strategies, as was mentioned above. Better, one might think, to develop an integrated data and Internet governance strategy given the extensive overlap and

---

[214] *See* Elinor Ostrom, *Polycentric Systems as One Approach for Solving Collective-Action Problems* 1–2 (Ind. Univ. Workshop in Pol. Theory and Pol'y Analysis, Working Paper Series No. 08–6, 2008). As originally explained by Professor Vincent Ostrom, "a polycentric political system would be composed of: (1) many autonomous units formally independent of one another, (2) choosing to act in ways that take account of others, (3) through processes of cooperation, competition, conflict, and conflict resolution." VINCENT OSTROM, THE MEANING OF FEDERALISM 225 (1991). The concept, though, has enjoyed wide application, including in the Internet governance context. *See* SCOTT J. SHACKELFORD, GOVERNING NEW FRONTIERS IN THE INFORMATION AGE: TOWARD CYBER PEACE (2020).

interests shared across these contexts. To help contextualize such an approach, aside from national security it is useful to consider what the end goals may be, and how best to build an international coalition to support them.

Finally, one must consider if AI strategies are a useful regulatory tool, especially considering the limitations briefly discussed. If AI strategies are to be used, more attention must be given to the complexity of the use of AI. Broader considerations of the complexity of AI must be considered and integrated into these strategies.

3.5 Opportunities for International Engagement

The foregoing analysis has highlighted several areas of convergence among nations as they strategize about future AI applications. In particular, as was discussed above, there is widespread agreement as to the importance of utilizing AI for public good, and human-centered design. One lens through which to view such efforts is sustainable development. Although there is not one universal definition of this concept, it is commonly understood as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs."[215] This definition has both public benefit and human-centered design at its heart. Increasingly, countries have been promoting an array of initiatives tied to sustainability, notably the UN Sustainable Development Goals. Indeed, the Sustainable Development Goals underscore the need for a broad conception of sustainability, one that includes environmental stewardship, reduced inequality, along with international peace and justice.[216] One 2020 study, for example, found that AI can "enable the accomplishment of 134 targets across all the [Sustainable Development] goals," but that it "may also inhibit 59 targets."[217] In other words, AI is a double-edged sword, which should be wielded with skill lest unintended consequences crowd out the potential positive benefits of AI on human rights and society.

CONCLUSION

[215] THE WORLD COMMISSION ON ENVIRONMENT AND DEVELOPMENT: OUR COMMON FUTURE 37 (1987). *See also* Gabcikovo-Nagymaros Project (Hung. v. Slovk.), 1997 I.C.J. 7, 78 (Sept. 25) (defining sustainable development as "[the] need to reconcile economic development with protection of the environment").

[216] *Sustainable Development Goals*, UN DEV. PROG., http://www.undp.org/content/undp/en/home/sustainable-development-goals.html (last visited Jan. 24, 2019)

[217] *See* Ricardo Vineusa, T*he Role of Artificial Intelligence in Achieving the Sustainable Development Goals*, Nature (Jan. 2020), https://www.nature.com/articles/s41467-019-14108-y.

This Article has analyzed the national AI strategies of dozens of nations across seven identified dimensions: transparency, accountability, security, privacy, fairness, human-centered design, and public benefit. We found that there was the greatest degree of convergence across these strategies within the domains of human-centered design and public benefit, in part given how ambiguous these terms may be. We suggested that sustainable development may be an overarching concept to help drive further international efforts in this arena that could be of particular interest to the Biden administration, and through which AI norm promotion may flourish such as in support of the Sustainable Development Goals.