AIWS REPORT ABOUT AI ETHICS

By Michael Dukakis, Nguyen Anh Tuan, Thomas Patterson, Thomas Creely, Nazli Choucri, Paul Nemitz, Derek Reveron, Hiroshi Ishiguro, Eliot Weinman, Kazuo Yano

Part I: Introduction

In 2018, the Michael Dukakis Institute for Leadership and Innovation (MDI) established the Artificial Intelligence World Society (AIWS) for the purpose of promoting ethical norms and practices in the development and use of AI. AI can be an important tool to serve and strengthen democracy, human rights, and the rule of law. Its misuse could undermine those ideals. In this effort, MDI invites participation and collaboration with think tanks, universities, non-profits, firms, and other entities that share its commitment to the constructive and development of AI.

One of MDI's AI-related initiatives is the Government AIWS Ethics and Practices Index, which is a measure of the degree to which governments engage in ethical AI practices.

This report describes the Index and its components. It is MDI's goal to get leading countries—those in the G7 and OECD, as well as other countries with populations exceeding 75 million—to apply the Index, which would contribute to the institutionalization of AI ethics and practices that serve the interests of humanity and do not threaten those interests.

The Index is in the draft stage and MDI invites suggestions on refining it. Once it's finalized, MDI will explore the possibility of an annual assessment of governments by evaluating their performance in the context of the Index.

The appendix to this report provides a cursory review of the strategies, activities, and progresses of selected countries in terms of ethical AI practices.

Part II: Overview of Government AIWS Ethics and Practices Index

The AIWS Ethics and Practices Index is designed to track the AI activities of governments. The Index measures the extent to which a government's AI activities respects human values and contributes to the constructive use of AI.

The Index has four components by which government performance is assessed:

1. **Transparency**. Substantially promotes and applies openness and transparency in the use and development of AI, including data sets, algorithms, intended impacts, goals, and purposes.
2. **Regulation**. Has laws and regulations that require government agencies to use AI responsibly; that are aimed at requiring private parties to use AI humanely and that restricts their ability to engage in harmful AI practices; and that prohibit the use of AI by government to disadvantage political opponents.
3. **Promotion**. Invests substantially in AI initiatives that promote shared human values; refrains from investing in harmful and exploitative uses of AI (e.g., autonomous weapons, propaganda creation and dissemination, social control).
4. **Implementation**: Seriously executes AI laws and regulations; respects and commits to widely accepted ethical principles and practices.

Governments would be assessed in each category by the standards of the moment. AI is in an early stage, and governments are only beginning to address the issue through, for example, laws and regulations. Later, as governments have had more time to assess the implications of AI, more substantial efforts would be expected—for example, a more fully articulated set of AI-related laws and regulations.

Part III: Particular Evaluative Criteria

AI can be a force for helping people achieve well-being and happiness, unleash their potential, obtain greater freedom, relieve them of resource constraints and arbitrary/inflexible rules and processes, and solve important issues, such as SDGs.

Among the criteria by which a government will be assessed in terms of its commitment to the positive use and development of AI are:

General

+ Whether its AI-related activities align with the positive purposes of AI.
+ Whether any of its AI-related activities have detrimental effects on well-being and happiness, restrict human potential; restrict human freedom; and impede solutions to society's problems.

Data sets.

+ Whether they disclose what data they collect, for what purposes, and from what sources; how they collect data (techniques & models).
+ Whether they have methods to encourage the proper use of data sets.
+ Whether they clearly communication and publicize governance programs or risk assessment tools to detect and remedy any possible discriminatory effects of the data and models used.

Algorithms.

+ Whether the collection, use, and management of data by AI algorithms follow ethical principles that promote fairness and avoid unjust effects on people, particularly in regard to characteristics such as race, ethnicity, gender, nationality, income, sexual orientation, ability, and political or religious belief.

Intended impacts

+ Whether they clearly communication intended impacts of AI development and use.
+ Whether they seek to prevent or mitigate adverse impacts and provide effective remediation of adverse impacts.
+ Whether the procure and deploy risk assessment tools.

Transparency in national resources:

+ Whether they publicize the identity of responsible departments for AI policy and project implementation.
+ Whether they publicize AI-related budgets and expenditures.

Refrain from harmful uses

+ Whether they refrain from AI design and deployment in areas such as: weapons or technologies that are likely to cause harm to humanity, that could cause injury to individuals, that violate internationally accepted norms and accepted principles of international law and human rights.

In addition to such considerations, governments should strive to ensure:

- *Responsibility for mistakes:* It must be clear where liability lies when systems make mistakes. General principles should guide accountability.
- *Transparency in decision making:* It must be clear when AI systems need to explain their actions to humans to show why a decision was made, and when, if ever, such transparency is not necessary.
- *Avoiding bias* - Steps must be made to stop systemic bias. Core values such as equality, diversity and lack of discrimination must be promoted.
- *Core ethical values* - What are the core ethical values of AI to be?
- *Data protection and Intellectual Property* - The importance of data protection, IP ownership and cyber security must be recognized and balanced against the need to use data to promote innovation.
- *Mitigating social dislocation* - The codes should confront what obligations rest on actors who deploy AI to mitigate the social dislocation that results.
- *Cybersecurity* - The need for strong protection against hacking will increase as AI systems take a heightened role in society.

Appendix: The current situation in selected countries

Governments are only beginning to address the issues related to AI through national strategies, laws and regulations. This part summarizes a few of the developments in selected countries.

1. The United States

Becoming a leader in the field of AI is one of top priorities of United States. The country also shows determination to pursue this goal through many specific actions and declarations.

On September 7, 2018, the U.S. Department of Defense announced that it will invest up to $2 billion over the next five years for the advancement of AI. This will be in addition to existing government spending on AI R&D, which totaled more than $2 billion in 2017.

To promote these activities, bills have been introduced in Congress that mention or focus on AI. There are at least nine bills that relate to autonomous driving, including The SELF DRIVE ACT introduced in the House on July 25, 2017. This bill establishes the federal role in ensuring the safety

of highly automated vehicles by encouraging the testing and deployment of such vehicles. The Department of Transportation must require safety assessment certifications for the development of a highly automated vehicle or an automated driving system.

Some bills relate to the economic impact of AI. The Innovation Corps Act of 2017, introduced in the House on March 16, 2017, requires the Department of Commerce to establish a competitive program to make grants to institutions of higher education to establish or enhance education programs that retrain workers displaced from their jobs by automation to provide such workers with skills needed for jobs in STEM fields. The AI JOBS Act, introduced in the House on January 18, 2018, requires the Department of Labor to prepare and submit to Congress a report on AI and its impact on the workforce.

Other bills address the technology and implications of AI itself, including The FUTURE of AI Act 2017 (introduced in both the House and Senate on December 12, 2017) which requires the Secretary of Commerce to establish a Federal Advisory Committee on the Development and Implementation of AI. The National Security Commission on Artificial Intelligence Act (introduced in the House on March 20, 2018) which considers the creation of an independent panel to explore advances in AI and assess their economic and national security impact, etc.

At the state and local levels, many AI-related bills have been introduced. On August 30, 2018, for example, the State of California unanimously adopted legislation in support of the Future of Life Institute's Asilomar AI Principles – a set of 23 principles intended to promote the safe and beneficial development of AI. Enacted on January 11, 2018, The New York Algorithm Monitoring Task Force of The New York City Council requires the creation of a task force that studies how city agencies use algorithms to make decisions that affect New Yorkers' lives. In December 2017, Supervisor David Canepa drafted a resolution calling for ban on development and use of autonomous weapons.

Steps towards AI policy in the U.S. were first taken by the Obama administration. The Obama White House published three globally influential reports, including "Preparing for the Future of Artificial Intelligence," "The National Artificial Intelligence Research and Development Strategic Plan," and "Artificial Intelligence, Automation, and the Economy." The first report made specific recommendations related to AI regulations, public R&D, automation, ethics and fairness, and security. The second report outlined a strategic plan for publicly funded R&D in AI. The final report examined in further detail the impact of automation and what policies are needed to increase the benefits of AI and mitigate its costs.

President Trump's White House has taken a free market-oriented approach to AI. He was the first U.S. president to specifically name AI as an Administration R&D priority in his 2019 Budget Request to Congress. The Trump administration seeks to remove regulatory barriers to innovation so that American companies have the flexibility to innovative and grow.

Michael Kratsios, Deputy Assistant to the President for Technology Policy, announced the U.S. Government has four goals: (1) maintain American leadership in AI; (2) support the American worker; (3) promote public R&D; and (4) remove barriers to innovation. To achieve these objectives, a new Select Committee on Artificial Intelligence was established to advise the White House on interagency AI R&D priorities and to consider the creation of federal partnerships with industry and academia. Another commission, National Security Commission on Artificial Intelligence, will also review ethical issues and national security risks associated with AI and machine learning technologies. Meanwhile, in July 2018, the Pentagon announced a new Joint AI Center that will have oversight over the majority of service and defense agency AI efforts.

2. Canada

The Canadian government launched the Pan-Canadian Artificial Intelligence Strategy in its 2017 Budget with the allocation of C$125 million to invest in AI research and talent in five years.

Being the first country to release a national AI strategy, Canada has four major goals in the AI strategy: (1) increase the number of AI researchers and graduates; (2) establish three clusters of scientific excellence; (3) develop thought leadership on the economic, ethical, policy, and legal implications of AI; (4) support the national research community on AI.

Canada's AI strategy is distinct from other strategies because it is primarily a research and talent strategy. It's initiatives—the new AI Institutes (including the Alberta Machine Intelligence Institute in Edmonton, the Vector Institute in Toronto, and MILA in Montreal), CIFAR Chairs in AI, and the National AI program—are all geared towards enhancing Canada's international profile as a leader in AI research and training.

3. Germany

Germany's national AI strategy is expected to be published at the Digital Summit 2018 in Nuremberg (December 3–4). In July 2018, Germany's federal cabinet released a paper "AI made in Germany" that outlines the goals of the national strategy to strengthen and expand German and European research in AI and focus on the transfer of research results to the private sector and the creation of AI applications. According to the paper, the German government has set aside around €3 billion for research and development of AI. Previously, basic research in AI in Germany has been funded by the German Research Foundation (DFG) with more than EUR 3 billion annually, while applied AI has been funded continuously by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung - BMBF), for a total of €215 million in the past thirty years (1988-2018).

The government pursues three key goals with the Artificial Intelligence Strategy: (1) Germany and Europe are to become leaders in the development and application of AI technologies. Germany's competitiveness is to be underpinned; (2) With the strategy, it is to be ensured that the development and use of AI is responsible and that it is pursued for the common good; (3) Within the framework of a broad dialogue within society and active political steering, AI is to be firmly anchored in society in ethical, legal, cultural and institutional terms.

The proposed plan is quite comprehensive and also includes measures to attract international talent, respond to the changing nature of work, integrate AI into government services, make public data more accessible, and promote the development of transparent and ethical AI.

In June 2017, the German Federal Ministry of Transport and Digital Infrastructure (BMVI) published ethical guidelines for self-driving cars in a report titled, "Ethics Commission: Automated and Connected Driving," which defines 20 ethical rules for automated and connected vehicular traffic.

Moreover, in September 2018, the Data Ethics Commission created by the Federal Government. The Commission has a mandate to develop recommendations for law and ethics of AI until October 2019.

4. The United Kingdom

With the intention for the UK to be the global leader in AI, the UK government announced it will invest over £300 million in private sector investment from domestic and foreign technology companies, expand the Alan Turing Institute, create Turing Fellowships, and especially launch the Centre for Data Ethics and Innovation.

In June 2017, The Parliament established the Select Committee on AI to further consider the economic, ethical and social implications of advances in artificial intelligence, and to make recommendations. In April 2018, the Committee published a report, "AI in the UK: ready, willing and able?" which considers AI development and governance in the UK. It acknowledges that the UK cannot compete with the US or China in terms of funding or people, but suggests the country may have a competitive advantage in considering the ethics of AI such as: (1) AI should be developed for the common good and benefit of humanity; (2) AI should operate on principles of intelligibility and fairness; (3) AI should not be used to diminish the data rights or privacy of individuals, families or communities; (4) all citizens have the right to be educated to enable them to flourish mentally, emotionally and economically alongside AI; (5) the autonomous power to hurt, destroy or deceive human beings should never be vested in AI.

Another organization, Centre for Data Ethics and Innovation, was established to review the existing governance landscape and advise the government on ethical, safe and innovative uses of data, including AI. And All Party Parliamentary Group on Artificial Intelligence is to address ethical issues, social impact, industry norms, and regulatory options for AI in Parliament.

The UK government also launched a Sector Deal for AI to advance the UK's ambitions in AI consistent with its Industrial Strategy, and taking into account the advice of the Parliament's Select Committee on AI.

5. France

"We have to be in a position to build, in France and in Europe, an artificial intelligence ecosystem," President Emmanuel Macron said in a speech. "We should have a policy of open data," he added, and "have to think on the subject from a political and ethical point of view … to come up with a common understanding and rules."

With an investment of nearly €1.5 billion euros for AI until 2020, President Macron aims to do nothing less than to lead his country into the age of AI and erase 30 years of underperformance on innovation. Details for the following have not be released, but €700 million will go towards research, €100 million this year to AI startups and companies, €70 million annually through France's Public Investment Bank, and €400 million to industrial projects in AI.

His plan consists of four components: (1) announce several initiatives to strengthen France's AI ecosystem and attract the international talent. Key among them was the announcement of the National Artificial Intelligence Programme, which will create a network of four or five research institutes across France; (2) develop an open data policy to drive the adoption and application of AI in sectors where France already has the potential for AI excellence, such as healthcare; (3) the government will create a regulatory and financial framework to support the development of domestic "AI champions;" (4) the government will development regulations for ethics to ensure that the use and development of AI is transparent, explainable, and non-discriminatory.

Another key aspect of President Macron's plan is to lure French AI researchers, many of whom occupy top positions in Silicon Valley, back to France.

France has also a national strategy for AI called "AI for Humanity," which is outlined in the "Villani Report" titled, "For a Meaningful Artificial Intelligence," which was unveiled in March 2018 with seven primary themes: (1) developing an aggressive data policy (to improve access to big data); (2) targeting four strategic sectors (healthcare, environment, transport, and defense), but President Macron only talked about the potential of AI for healthcare and transportation; (3) boosting the potential of French research (and investing in talent); (4) planning for the impact of AI on labour; (5) making AI more environmentally friendly; (6) opening up the black boxes of AI; (7) ensuring that AI supports inclusivity and diversity.

Another algorithm-related report, titled, "HOW CAN HUMANS KEEP THE UPPER HAND? The ethical matters raised by algorithms and artificial intelligence," highlighted two founding principles to "put AI at the service of humans," including "A principle of loyalty" and "A principle of continued attention and vigilance." The report also made six policy recommendations involving

education about ethics, improving interpretability and auditability of AI systems, and promoting human freedom.

6. Italy

In March 2018, Italy published a white paper titled, "Artificial Intelligence: At the service of citizens," which focuses on research and development or private sector uptake. It recommends about how to develop better public services with the use of AI that could help eliminate inequalities and measure impact, and describes ethical, legal, technological, and cultural challenges.

The white paper was created by an AI Task Force, launched in September 2017 by the Agency for Digital Italy. The goals of the AI Task Force are to analyze the use of AI in public administration to enhance the quality of services offered to citizens, define the potential of AI services and applications, and explore the ethical and social implications of the use of AI.

Moreover, in July 2018, a consortium of universities and research centers in Italy united to create a new national laboratory for AI, National Laboratory of Artificial Intelligence and Intelligent Systems (CINI-AIIS Lab), which aims to contribute to ongoing national scientific and technological initiatives such as the National Clusters and Competence Centers of the Enterprise 4.0 program, as well as to industrial research and Public Administration activities.

7. Japan

Japan developed a national AI strategy Artificial Intelligence Technology Strategy and released it in March, 2017. Japan is planning to increase its science and innovation budget by JPY900 billion by 2020 for AI. Different Japanese ministries are also funding R&D centers, e.g., Ministry of Economy, Trade and Industry (METI) is funding R&D center at the National Institute of Advanced Industrial Science and Technology (AIST). The strategy shows Japan's Industrialization Roadmap, which envisions AI as a service and organizes the development of AI into three phases: The strategy applies this framework to three priority areas of Japan's Society 5.0 Initiative— productivity, health, and mobility—and outlines policies to realize the industrialization roadmap. In June 2018, the Japanese government announced that artificial intelligence would also become an official part of its "integrated innovation strategy."

Based on instructions issued by the Prime Minister in "Public-Private Dialogue towards Investment for the Future" in April 2016, the national government 3 established the "Strategic Council for AI Technology". The Council, acting as a control tower, has come to manage 5 National Research and Development Agencies that fall under the jurisdiction of the Ministry of Internal Affairs and Communications, Ministry of Education, Culture, Sports, Science and Technology, and Ministry of Economy, Trade and Industry:

On July 28, 2017, Japan published Draft AI R&D GUIDELINES for International Discussions in preparation for the Conference toward AI Network Society. Principles mainly concerning mitigation of risks associated with AI systems with 9 principles: Principle of collaboration, Principle of transparency, Principle of controllability, Principle of safety, Principle of security, Principle of privacy, Principle of ethics (respect human dignity and individual autonomy), Principle of user assistance, Principle of accountability.

8. China

China has announced its ambition to become a world leader in A.I. by 2030 in the national strategy A Next Generation Artificial Intelligence Development Plan divided in 3 steps (By 2020: reach globally advanced levels, By 2025: Achieve some world-leading breakthroughs, By 2030: become the world leader in AI by 2030). The report outlines China's strategy to build a domestic AI industry worth nearly US$150 billion, with four major tasks: (1) focus on developing intelligent and networked products such as vehicles, service robots, and identification systems; (2) emphasize the development AI's support system, including intelligent sensors and neural network chips; (3) encourage the development of intelligent manufacturing; (4) improve the environment for the development of AI by investing in industry training resources, standard testing, and cybersecurity.

In China, governments play a deliberate and explicit role in funding scientific research (giving US$800,000 to US$1million in subsidies to AI companies). Individual ministries and departments are responsible for implementing AI solutions across different sectors. China's government has also partnered with large private companies including Baidu and Tencent to undertake fundamental and applied research across different AI topics, e.g., Baidu is working with the Chinese government to develop brain-inspired intelligent technology.

In its national strategy report, China also shows plan to establish of artificial intelligence laws and regulations, ethical norms and policy systems, and form artificial intelligence safety assessment and control capabilities; standards and the intellectual property system for AI technology; safety supervision and evaluation systems for AI. "One group of Chinese leader is increasingly engaged with issues of AI safety and ethics. A wide range of Chinese AI researchers are also involved with translating the IEEE's Ethically Aligned Design report, as part of the Global Initiative for Ethical Considerations in Artificial Intelligence. However, some Chinese AI leaders dismiss calls for regulation and philosophizing." (Deciphering China's AI Dream)

The Chinese Government proposed in 2014 an extensive data collection system called the Social Credit System, which will monitor every action Chinese citizen perform by collecting vast amounts of data online, and then provide a score to measure how trustworthy each individual is, based on political, social, commercial and legal credit. If all goes well in the current planning and test phase

the system will be fully implemented by 2020 and will enlist every citizen living and every company operating in China. This proposal of China raises various ethical and societal questions.

9. Russia

Though Russia has not had an official national strategy, the Ministry of Education and Science of the Russian Federation, or MES; and the Russian Academy of Sciences released an ambitious 10-point plan in March, 2018, which recommends steps for prospective development of artificial intelligence in the country's federal and academic bodies. With assertion of President Vladimir Putin that "whoever becomes the leader in this sphere [AI] will become the ruler of the world", the Russian government is increasingly developing and funding various AI-related projects, many under the auspices of the Ministry of Defense and its affiliated institutions and research centers. At the moment, Russia's annual domestic investment in AI is estimated probably around 700 million rubles ($12.5 million) – a paltry sum next to the billions being spent by American and Chinese companies.

Russia has advocated for military uses of AI and been dismissive of ongoing deliberations at the United Nations to prohibit lethal autonomous weapon systems. Moreover, President Trump's National Security Strategy of the U.S in December, 2017 mentioned the problem that "Russia uses information operations as part of its offensive cyber efforts to influence public opinion across the globe. Its influence campaigns blend covert intelligence operations and false online personas with state-funded media, third-party intermediaries, and paid social media users or 'trolls'."

10. India

In June 2018, the Indian government released "National Strategy for Artificial Intelligence #AIforAll." The NITI Aayog paper identifies its national purpose: (1) enhance and empower Indians with the skills to find quality jobs; (2) invest in research and sectors that can maximize economic growth and social impact; and (3) scale Indian-made AI solutions to the rest of the developing world. Overall, the strategy determines to establish India as an "AI Garage" for emerging and developing economies.

NITI Aayog, the government think tank provides over 30 policy recommendations to invest in scientific research, encourage reskilling and training, accelerate the adoption of AI across the value chain, and promote ethics, privacy, and security in AI. Its flagship initiative is a two-tiered integrated strategy to boost research in AI. First, new Centre of Research Excellence in AI (COREs) will focus on fundamental research. Second, the COREs will act as technology feeders for the International Centre for Transformational AI (ICTAIs), which will focus on creating AI-based applications in domains of societal importance.

About the problem of ethics, the report recommends setting up a consortium of Ethics Councils at each CORE and ICTAI, developing sector specific guidelines on privacy, security, and ethics.

In the Union budget 2018, the Indian government allocated $480 million to fifth generation technology startups like artificial intelligence, machine learning (ML), Internet of Things (IoT), 3D printing and blockchain, committed to the vision and mission of making 'Digital India' a reality.

11. The European Union

The European Union is a political and economic union of 28 member states that are located primarily in Europe and has the European Commission in Brussels as Executive and body of political initiative.

In April 2018, the European Commission presented a Strategy Communication for AI in Europe, known "The European AI Alliance." In December, the European Commission will also publish a related action plan.

Ethics and legal regulation of AI is at the core of the European Union Strategy, and one key part of law effectively regulation AI is the General Data Protection Regulation which applies whenever AI works with personal data. It gives people far going rights to maintain control of their personal data, such as a right to deletion, and far going information rights, such as a right to understand programs such as AI which are treating their personal data and how these programs reach decisions, in case of automated decision making. They also have the right under the EU General Data Protection Regulation (GDPR) to object to automated decisions and require human intervention.

References:

Tim Dutton. An Overview of National AI Strategies. 2018 [https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd]

The European Commission. The European AI Alliance. 2018 [https://ec.europa.eu/digital-single-market/en/european-ai-alliance]

The Future of Life Institute. National and International AI Strategies. 2018 [https://futureoflife.org/national-international-ai-strategies/?cn-reloaded=1]