REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
09-07-2022	Journal Article	2018-2023

4. TITLE AND SUBTITLE

Framework for Global Accord on Artificial Intelligence (AI)

5a. CONTRACT NUMBER

5b. GRANT NUMBER

H98230-18-D-00-0010

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Nazli Choucri, Massachusetts Institute of Technology

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Massachusetts Institute of Technology
77 Massachusetts Avenue,
Cambridge, MA, 02139.

8. PERFORMING ORGANIZATION REPORT NUMBER

https://hdl.handle.net/1721.1/141737

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Science of Security & Privacy Program
National Security Agency,
US Department of Defense.

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

Choucri, N. (2022). Framework for global accord on artificial intelligence (AI) (Working Paper No. 2022-8). MIT Political Science Department. https://www.ssrn.com/abstract=4174485.

14. ABSTRACT

Advances in information and communication technologies – global Internet, social media, Internet of Things, and a range of related science-driven innovations and generative and emergent technologies – continue to shape a dynamic communication and information ecosystem for which there is no precedent.

These advances are powerful in many ways. Foremost among these in terms of salience, ubiquity, pervasiveness, and expansion in scale and scope is the broad area of artificial intelligence. They have created a new global ecology; yet they remain opaque and must be better understood— an ecology of "knowns" that is evolving in ways that remain largely "unknown." Especially compelling is the acceleration of Artificial Intelligence – in all its forms – with far-ranging applications shaping a new global ecosystem for which there is no precedent.

This chapter presents a brief view of the most pressing challenges, articulates the logic for worldwide agreement to retain the rule of law in the international system, and presents salient features of an emergent International Accord on Artificial Intelligence. The Framework for Artificial Intel

15. SUBJECT TERMS

Artificial intelligence, global accord, framework, ethics, rights, responsibilities.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Nazli Choucri
Unclassified	Unclassified	Unclassified	Same as Report	14	19b. TELEPHONE NUMBER *(Include area code)* 617-253-6198

Framework for Global Accord on Artificial Intelligence (AI)

Nazli Choucri

Professor of Political Science
Massachusetts Institute of Technology

July 9, 2022

Abstract

Advances in information and communication technologies – global Internet, social media, Internet of Things, and a range of related science-driven innovations and generative and emergent technologies – continue to shape a dynamic communication and information ecosystem for which there is no precedent.

These advances are powerful in many ways. Foremost among these in terms of salience, ubiquity, pervasiveness, and expansion in scale and scope is the broad area of artificial intelligence. They have created a new global ecology; yet they remain opaque and must be better understood—an ecology of "knowns" that is evolving in ways that remain largely "unknown." Especially compelling is the acceleration of Artificial Intelligence – in all its forms – with far-ranging applications shaping a new global ecosystem for which there is no precedent.

This chapter presents a brief view of the most pressing challenges, articulates the logic for worldwide agreement to retain the rule of law in the international system, and presents salient features of an emergent International Accord on Artificial Intelligence. The Framework for Artificial Intelligence International Accord (AIIA) is an initial response to this critical gap in the system of international rules and regulations.

Keywords

Artificial intelligence, global accord, framework, ethics, rights, responsibilities.

Citation: Choucri, N. (2022). *Framework for global accord on artificial intelligence (AI)* (Working Paper No. 2022-8). MIT Political Science Department.

Unique Resource Identifier: https://www.ssrn.com/abstract=4174485

Version: Author's final manuscript.

Is Version of: Choucri, N. (2021). Framework for an AI international accord. In Nguyen Anh Tuan (Ed.), *Remaking the World: Toward an Age of Global Enlightenment* (pp 27–43). Boston Global Forum/United Nations Academic Impact. ISBN: 978-0-578-93779-3. https://un100.net/books-un/the-book-remaking-the-world-the-age-of-global-enlightenment-available-on-july-1-2021/

Table of Contents

Framework for Global Accord on Artificial Intelligence

1 Emergent Global Challenges

Advances in information and communication technologies – global Internet, social media, Internet of Things, and a range of related science-driven innovations and generative and emergent technologies – continue to shape a dynamic communication and information ecosystem for which there is no precedent.

These advances are powerful in many ways. Foremost among these in terms of salience, ubiquity, pervasiveness, and expansion in scale and scope is the broad area of artificial intelligence. They have created a new global ecology; yet they remain opaque and must be better understood—an ecology of "knowns" that is evolving in ways that remain largely "unknown." Especially compelling is the acceleration of Artificial Intelligence – in all its forms – with far-ranging applications shaping a new global ecosystem for which there is no precedent.

This chapter presents a brief view of the most pressing challenges, articulates the logic for worldwide agreement to retain the rule of law in the international system, and presents salient features of an emergent International Accord on Artificial Intelligence. The Framework for Artificial Intelligence International Accord (AIIA) is an initial response to this critical gap in the system of international rules and regulations.

2 New Reality - New "Unknowns"

The term "artificial intelligence" refers to the theory and development of computer systems able to perform tasks that normally require human intelligence—such as visual perception, speech recognition, decision-making, translation between languages, self-driving cars, and so forth. It also spans efforts to establish machine-human brain connectivity in ways that are highly exploratory and whose implications are yet to be articulated to any great extent.

We are at the beginning of a new era, a world of mind-machine convergence. Its current logic, situated at the frontiers of biological intelligence and machine intelligence, is generally anchored in past data and has made possible whole new sources and forms of design space. Fully understanding the scale of the AI domain remains elusive. We have seen a shift from executing instructions by humans to replicating humans, outperforming humans, and transcending humans.

Almost everyone appreciates that advances in AI have already altered conventional ways of viewing and managing the world around us. We have created new realities for everyone—as well

as new possibilities. Nonetheless, when all is said and done, the “intelligence” that is “artificial” remains devoid of autonomous consciousness, empathy, and perhaps select other human features – such as ethics –so fundamental to humanity and the social order. It goes without saying that, sooner or later, humans will program machines to generate what we consider consciousness to be. Already we are seeing major efforts and assessments in that direction.

The expansion of Artificial Intelligence is widely recognized and could change our lives in ways yet unimagined. This expansion has created a new global ecology, one that remains opaque and poorly understood.

3 Call for Accord on Artificial Intelligence

The world of artificial intelligence today is framed by a set of unknowns – known unknowns and unknown unknowns – where technological innovations interact with the potential for the total loss of human control. Especially elusive is the management of embedded insecurities in applications of this new, ubiquitous technology and the imperatives of safety and sustainability.

But without adequate guidelines and useful directives, the undisciplined use of AI poses risks to the wellbeing of individuals and creates fertile ground for economic, political, social, and criminal exploitation. The international community recognizes the challenges and opportunities, as well as the problems and perils, associated with AI. Many countries have already announced national strategies to promote the proper use and development of AI. At the operational level, there are as yet no authoritative modes and methods for reviewing and regulating algorithms. This is yet another “open” space, in the full sense of the word.

We are now faced with a critical imperative, namely, to address head-on the policy issues raised by AI advances and to assess, evaluate, and respond effectively. We must engage in serious dialogue – buttressed by tolerance, learning, and mutual understanding – to converge on principles and practices of an agreement among members of the global society on a strategy to generate and enhance social benefits and wellbeing for all, shared by all.

At the core of this imperative is to establish a common understanding of policy and practice, anchored in general principles to help maximize the "good" and minimize the "bad" associated with AI. Given such ambiguities and uncertainties, it is not surprising that the international community has not yet fully grasped the implications of the new “unknowns” and the potential threats to the global order.

While individually, as well as jointly, these new capacities transcend the prevailing frameworks for maintaining order – nationally or internationally – on balance, the overall patterns appear not to generate a semblance of order. Some countries have developed national policies for the cyber domain, most notably regarding cybersecurity, Others may be in the process. Different

countries may impose different measures, individually or collectively, but for the most part new innovations and novel applications remain largely unregulated.

We have created new tradeoffs that must be assessed. We must now focus on critical principles and essential supporting practices for the new and emerging world that we have created. We must also envisage fundamental “best practices” for realities that have yet to emerge.

3.1 Toward a Worldwide Consensus

We must now re-think and consolidate the best practices for human development, recognizing the power and value of the individual and of society. Much yet is to be done.

An added factor is that AI is also becoming a focus for foreign policy and international cooperation. There is a shared view that no country will be able to compete or meet the needs of its citizens without increasing its AI capacity. At the same time, many countries are now engaged in technology leapfrogging. It is no longer expected, nor necessary, to replicate the stages of economic development of the West—one phase at a time. While the possibilities are varied and diverse, there is also is a clear awareness of the challenges and opportunities, as well as the problems and perils of and many are seeking ways of managing their approach to AI. At least 20 countries have announced formal strategies to promote the use and development of AI.

No two strategies are alike; however, there are common themes even among countries who focus on different aspects of AI policy. The most common themes addressed include those pertaining to scientific research, talent development, skills formation, public and private collaboration, visualization for innovation, and data standards and regulation, among others.

Transcending the diversity of situations and orientations, there is a solid foundation of shared goals in the international community, buttressed by the activities of United Nations agencies to facilitate operational strategies and assist in implementation of objectives. These include a general appreciation of skills, education, and talent development, public and private policy innovation, attention to fairness, transparency, and accountability, ethics and values of inclusion, reliability, security and privacy, science-policy links, standards for regulations and data development, and digital infrastructure.

In sum, all countries are going through a common experience of adapting to and managing unknowns. All of these venues are generally framed within an overarching context of sustainable development. All of this creates an international atmosphere welcoming of an International Accord on Artificial Intelligence on a global scale.

3.2 Logic for AI International Accord

There is a long tradition of consensus-based social order founded on cohesion and agreement, and not the use of force nor formal regulation or legislation. It is often a necessary precursor for managing change and responding to societal needs.

The foundational logic addresses four premises: What, Why, Who and How?

3.2.1 What?

An international agreement on AI is about supporting a course of action that is inclusive and equitable. It is designed to focus on relationships among people, governments, and other key entities in society.

3.2.2 Why?

To articulate prevailing concerns and find common convergence. To frame ways of addressing and managing potential threats – in fair and equitable ways.

3.2.3 Who?

In today's world, framing an international accord for AI must be inclusive of:

- Individuals as citizens and members of a community
- Governments who execute citizen goals
- Corporate and private entities with business rights and responsibilities
- Civil society that transcends the above
- AI innovators and related technologies, and
- Analysts of ethics and responsibility.

None of the above can be "left out." Each of these constitutes a distinct center of power and influence, and each has rights and responsibilities.

3.2.4 How?

The starting point for implementation consists of five basic principles to provide solid anchors for Artificial Intelligence International Accord.

1. *Fairness and Justice for All*: The first principle is already agreed upon in the international community as a powerful aspiration. It is the expectation of all entities – private and public – to treat, and be treated, with fairness and justice.

2. *Responsibility and accountability for policy and decision—private and public*: The second principle recognizes the power of the new global ecology that will increasingly span all entities worldwide—private and public, developing and developed.
3. *Precautionary principle for innovations and applications*: The third principle is well established internationally. It does not impede innovation but supports it. It does not push for regulation but supports initiatives to explore the unknown with care and caution.
4. *Ethics-in-AI*: Fourth is the principle of ethical integrity—for the present and the future. Different cultures and countries may have different ethical systems, but everyone, everywhere recognizes and adopts some basic ethical precepts. At issue is incorporating the commonalities into a global ethical system for all phases, innovations, and manifestations of artificial intelligence.

Jointly, these four foundations – What, Why, Who, How – create powerful foundations for framing and implementing an emergent Artificial Intelligence International Agreement.

4 Artificial Intelligence Global Accord

The AIIA Draft Framework recognizes pathbreaking initiatives – notably the Budapest Convention on Cybercrime and the European Union General Directive – that signal specific policies to protect the integrity of information and the values that support this integrity. In addition, it recognizes the ongoing deliberations in the European Union regarding the future of AI and best means of supporting EU objectives, as well as those of member states. Then, too, the Draft Framework acknowledges the deliberations of the United States National Commission on Artificial Intelligence, and the Report of its results.

Consistent with the legal principle of a rules-based international community, the Draft Framework consists of several initial procedural and operational strategies, as follows:

(1) Preamble to highlight critical values and conditions to help clarify the underlying commonalities among all signatory entities supporting an AIIA of worldwide scale and scope.

(2) Framework Design to define the parameters of structure and process for further global deliberation and refinement.

(3) Operational Measures to buttress pragmatic as well as aspirational purposes, and Support System for realizing, formalizing and implementing an International as well as Global and International Accord on Artificial Intelligence.

Each calls for some articulation.

4.1 Preamble

The Preamble to the AIIA Framework is predicated on critical premises that reflect important features of the results-based system that defines today's international community, and are assumed to be operative at the drafting of the Framework. These are assumptions that enable framing of further order, and are stated as follows:

- Recognizing accelerated innovations in and applications of AI in diverse facets of the human condition. All advances and applications thereof must be coupled with, and adhere to, the internationally recognized precautionary principle.
- Supporting the international community's commitment to human rights. The potential harms on society inflicted by unrestrained uses of AI must be prevented in all contexts and situations everywhere.
- Convinced of the salience of rights, commensurate attention must be given to responsibilities.
- Understanding the differences and discrepancies among countries in computational skills and innovations in AI, a worldwide AI educational initiative must be designed to enable "full recognition" of all challenges surrounding AI.
- Respecting the diversity of the international community, all parties, public and private, all measures for implementation will be taken by national authorities.
- Acknowledging that that the dearth of guidelines may evolve into chaotic and undisciplined conditions that undermine benefits of AI to society by enabling exploitation and damage.

4.2 Framework Design

Consistent with the principles the provisions of the Budapest Convention on Cybercrime as well as the EU General Directives, and respecting the Social Contract for the AI Age, the AIIA draft framework is conceived and designed as:

- A multi-stakeholder, consensus-based international agreement to establish common policy and practice in development, use, implementation and applications of AI.
- Anchored in the balance of influence and responsibility among governments, businesses, civil society, individuals, and other entities2
- Respectful of national authority and international commitments and requires assurances of rights and responsibilities for all participants and decision-entities.

To consolidate the design into a formal International Accord, it is essential to:

- Review legal frameworks for AI at various levels of aggregation to identify elements essential for an international AI legal framework.
- Recognize methods to prevent abuses by governments and businesses in uses of AI, Data, Digital Technology, and Cyberspace (including attacks on companies, organizations, and individuals, and other venues of the Internet).
- Consolidate working norms to manage all aspects of AI innovations, and
- Construct and enable response-systems for violations of rights and responsibilities associated with the development, design, applications, or implementation of AI.

5 Process and Essential Measures

Given that "unrestricted use" of AI is not deemed acceptable by the international community, and a "total ban" may be unreasonable at this point, the Draft Framework for AIIA puts forth a set of measures for immediate review, assessment, refinement, and adoption by the international community. These measures are addressed to actors and entities]

5.1 Individual Rights and Responsibilities

The scope of rights includes:

- Rights pertaining to Data and the Internet.
- Rights to digital and AI related education.
- Rights to political participation in AI policy deliberations.
- Right to avoid digital damages.

And with rights, come responsibilities.

- Avoid digital damages
- Contribute to the common good
- Participate in codes of digital ethics
- Remain cognizant of AI applications
- Refrain from use of malware or distribution of misinformation

5.2 Imperatives for National Policy

Governments are required to:

- Implement the AI governance policies, standards, and norms adopted by the international community.
- Provide education for all citizens – “real” or online – with advanced AI Technology.
- Design incentives and directives for responsible AI use.
- Protect intellectual property rights without undermining free access to the information commons.

5.3 Collaborations among States

Inter-state collaboration is required to:

- Support shared AI policies and common goals
- Enable international measures by creating national policies and instruments
- Reinforce protection of human rights in AI innovations and uses.
- Develop common principles and methods to contain and combat misinformation
- Recognize the Social Contract for the AI Age.
- Establish a Worldwide Alliance toward Digital Governance

5.4 United Nations and International Organizations

These entities are expected to:

- Enable and support sustained data collection and analysis
- Provide guidelines for worldwide AI knowledge and education
- Create support systems for global digital ethics
- Establish international legal foundations for management of Ak d Convene all willing entities to participate in the framing and forging international judicial systems devoted to AI applications
- Contribute to the United Nations Centennial, notably a Global Enlightenment Prize and international Lecture
- Reinforce the AIWS City initiatives to develop and evaluate a People Centered-Economy.

5.5 Business Entities

National and international business are expected to>

- Enable independent audits for transparency, fairness, accountability, and cybersecurity.
- Adopt common AI values, standards, norms, and data ownership rules with penalties for noncompliance.
- Collaborate with governments, civil society, and international organizations to help create a people-centered AI, data, and Internet ecosystem
- Support sanction-systems to enforce a rules-based international order.

5.6 Civil Society

Rights and responsibilities of civil society organizations include>

- Monitoring governments and firms in support of common values and standards.
- Enabling all forms of voluntary data, analytics and other cooperatives, including the pooling by individuals of their personal data for the benefit of the group or community, conforming to international norms.
- Supporting an intelligent, thoughtful development and use of knowledge, as well as institutional opportunities for knowledge.

6 The Support System for AIIA Framework

Based on the internationally recognized Precautionary Principle, the support system for AIIA Framework is expected to facilitate and formalize the Framework and its implementation. The supports include the following products and processes:

- Code of Ethics for AI Developers and AI Users.
- Operational systems to monitor AI performance by governments, companies, and individuals.
- Certification for AI Assistants to enable compliance to new standards.
- Establish a multidisciplinary scientific committee to provide independent review and assessment of innovations in AI and directives for safe and secure application, consistent with human rights and other obligations.
- Enable a Social Contract for the AI Age to be supported by United Nations, governments, companies, civil society and the international community.
- Consolidate World Alliance for Digital Governance as the global authority to enforce the emergent accord.

7 End Note: Challenges, Opportunities, Next Steps

This End Note addresses briefly some salient challenges, followed by highlights of opportunities, and concludes with a brief word of caution.

7.1 The Challenges

Technology and innovation are growing much faster than the regulatory framework anywhere, and most certainly at the international levels. Of course, we do not want regulations to change at the level of technological change – that would create chaos; you can imagine why and how.

We can expect innovations in AI to grow much faster than has been the case so far – due in large part to new generations being educated in AI early on. We tend to think that the key players are in the AI arena are companies, governments, and academic researchers. We are overlooking youth as the growth-asset that will buttress both society and AI in the decades to come. It is foolhardy to ignore what are likely to be the real challenges, namely, the scale and scope of (a) unknowns, and (b) unknown "unknowns," and (c) their intended and unintended consequences, individually and collectively.

7.2 The Opportunities

The international community has a long and effective record of framing and reaching agreement in almost all areas of interaction. These are especially powerful in areas of standards, quality controls, certifications and so forth. As a result, we should take stock of what we do know about what works best in different areas and domains.

Furthermore, how and why do we know what works best? These questions are designed to empower researchers, businesses, government agencies, and international entities – private and public – to address how and why? Then, too, given the known "unknowns," what should we know? We have an opportunity to mine our own records for the "best fit" with the properties and conditions surrounding the current Artificial Intelligence dilemma.

Among the major opportunities before us is to inquire: What is the best precedent? Is it nuclear power? Is it climate change? What are other high-risk areas? Usually, we respond to such questions long after the fact. But can we avoid this delay? At this point, we have an opportunity to seriously consider the properties of a global accord in AI before we are faced with a major disaster.

Of high value, for example, is to consider and address the role of ethics in courses on innovations in AI, as well as ethics for all uses and users. So, too, it is important to focus on international law relevant to AI. There are many other high-value issues to consider at this point. The reason is this: The lines of political contention are not yet clearly drawn among potentially conflicting perspectives (or countries).

Therefore, now is the opportunity to proceed before these are consolidated into “lines in the sand.”

7.3 The Next Steps

At this point governments do not control AI innovation and/or diffusion. Much of the action taken is not in the public sector. Individuals and non-state groups matter and matter a lot. Constituencies are varied and overlapping. Consensus building is essential for society, not only for governments. Any position taken must be in the interest of everyone. Any initiative cannot be seen to dampen innovation or markets.

At the same time, we know from experience that “punishments”—in their various forms, do not work. We are in a world where large firms in the IT and communication business area are very powerful. Many are larger than most countries. The dilemma becomes: Whom to punish?

The immediate next step is establishing a multi-stakeholder support base. This is a necessary step to get to the point of articulation of interests and negotiations for “best outcomes.”

We are now dealing with 21st century realities wherein state coalition building is essential. We could even initiate a global competition among young minds for creating the best international agreement on artificial intelligence.