



Knowledge across boundaries: Promoting global cooperation on AI regulation

ART-AI submission to the Global Digital Compact

Theme 6: Promoting regulation of artificial intelligence

Supplementary Report



art-ai



UNIVERSITY OF
BATH

Authors:

Report prepared by Catriona Gray* Huixin Zhong, Joseph Marvin Imperial, Jack McKinlay and Eoin Cremen following collective discussions with Charles Larkin, Marina De Vos, Emma Carmel, Iulia Cioroianu, Julian Padget, and Elena Safrygina.

With thanks to Janina Hoffman for her constructive feedback on an earlier draft.

*corresponding author cg909@bath.ac.uk

Published: March 2023



Knowledge across boundaries: Promoting global cooperation on AI regulation by Gray, Catriona et al. is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

Contents

Background.....	4
Introduction.....	5
Box 1: Summary of Principles and Recommendations.....	6
1. The complex nature of AI regulation: common dilemmas	7
Dilemma 1: Definitions.....	7
Box 2: Interdisciplinarity	8
Dilemma 2: Purpose	8
Dilemma 3: Sectoral shape.....	9
Box 3: The healthcare sector	10
Box 4: The public sector	11
Dilemma 4: Prescriptiveness	11
Dilemma 5: Extra-regulatory tools	12
Box 5: Auditing AI.....	13
Box 6: Responsible AI Licences (RAILs).....	14
2. The global nature of AI regulation: the role of the UN	15
Box 7: Lethal Autonomous Weapons Systems	16
3. Principles and Recommendations	17
PRINCIPLES.....	17
RECOMMENDATIONS	18

Background

In September 2021, the Secretary-General of the United Nations, António Guterres, released the report [Our Common Agenda](#). It made broad proposals for the future of multilateralism, which built on the 2020 Roadmap for Digital Cooperation. It proposed a Global Digital Compact to be agreed at the Summit of the Future, which will be held in 2024. In October 2022, the President of the UN General Assembly appointed the Permanent Representatives of Rwanda and of Sweden as Co-facilitators to lead the intergovernmental process

The [Roadmap for Digital Cooperation](#) was published in 2020 and set out key actions, including “Supporting global cooperation on Artificial Intelligence that is trustworthy, human rights based, safe and sustainable and promotes peace.”

The Global Digital Compact will be the main outcome of a multistakeholder technology track and is expected to outline shared principles for an open, free and secure digital future for all. The Common Agenda report set out themes that the Global Digital Compact is likely to cover:

- Connect all people to the internet, including all schools;
- Avoid internet fragmentation;
- Protect data;
- Apply human rights online;
- Introduce accountability criteria for discrimination and misleading content;
- Promote regulation of artificial intelligence; and
- Digital commons as a global public good.

The following submission to this consultation focuses on the theme of promoting regulation of artificial intelligence.

The Global Digital Compact consultation invited submissions outlining suggested Principles and Recommendations. This supplementary report sets out our proposed Principles and Recommendations (in section 3) and provides greater detail and justification for these proposals (in sections 1 and 2).

Introduction

We are a group of academic researchers affiliated to ART-AI, a centre for doctoral training in **accountable, responsible, and transparent AI** at the University of Bath, UK. Our expertise spans multiple disciplines including economics, computer science, law, public policy, psychology, health, sociology, and electronic engineering. We draw on our **collective knowledge and experience** as AI researchers and practitioners.

High-level principles – including those of accountability, responsibility, and transparency – are now widely affirmed in instruments such as the UNESCO Recommendation on the Ethics of Artificial Intelligence. These principles alone, however, do not and cannot offer adequate solutions to all regulatory concerns. Even with these principles in mind, policymakers attempting to formulate and enact effective, coherent, and legitimate AI regulation encounter **several common dilemmas**. These include issues relating to:

1. **Definitions:** What is AI? How do we determine which systems fall within this category? Should AI be defined in law?
2. **Purpose:** What are the main rationales for regulation – e.g., to protect fundamental rights, to integrate markets, or to drive growth?
3. **Sectoral shape:** Should AI regulation be designed to apply uniformly across all sectors, or adjusted to fit specific sectoral contexts?
4. **Prescriptiveness:** How obligatory should regulatory interventions be? How precise and elaborated?
5. **Extra-regulatory tools:** What is the appropriate role for tools such as technical standards, AI audits, licences, and transparency registers in the regulatory mix?

We do not attempt to resolve these dilemmas in this report. Our objective is instead to highlight the **complex and global nature of AI regulatory policymaking**,¹ and to demonstrate the need for **new and hybrid approaches to knowledge** that recognise this.

Many of the challenges associated with the development and adoption of **AI technologies are global in character** and directly within the mandates of existing UN bodies. Regulation of AI is increasingly relevant to fields including peace and security, migration governance, humanitarian protection, and climate action. States alone are not well placed to regulate all problems arising from the use of AI. Without careful consideration of the global dimensions of AI regulation, there is a risk of fragmentation and a race to the bottom in standards. Without global cooperation, governments will be less able to address growing inequalities in wealth, power, and access to new technologies². The UN itself has an important role to play in **facilitating global cooperation on AI regulation**.

¹ Smuha, N. A., 2021, [From a 'race to AI' to a 'race to AI regulation': regulatory competition for artificial intelligence](#).

² Gray, C., 2023, [More than Extraction: Rethinking Data's Colonial Political Economy](#).

Developments in AI, including its widespread adoption, pose some of the most complex, transversal, and controversial policy questions of our time. In response, we propose that **interdisciplinary academic research** play a central role in guiding policymakers and regulators. **Open dialogue** and **knowledge exchange between researchers, policy actors, and indeed citizens**, will be critical to advancing the Global Digital Compact’s vision of an open, free, and secure digital future for all.

There is growing consensus around normative principles to guide AI policymaking. With this in mind, we have chosen to focus our proposed Principles and Recommendations on issues of **knowledge production, exchange, and translation**, and on the unique role that the **UN itself can play** in promoting cooperation on AI regulation.

The Secretary-General's 2020 Roadmap for Digital Cooperation recommended action to support global cooperation on AI. The Roadmap set out a vision for AI that is trustworthy, human rights based, safe and sustainable, and that promotes peace. To achieve this goal, we strongly believe there is a need to promote research that is **independent, interdisciplinary, and international**. We propose four principles to guide these efforts, along with four recommendations for the Global Digital Compact. These Principles and Recommendations are elaborated further in section 3.

Box 1: Summary of Principles and Recommendations

Principles	Recommendations
<ol style="list-style-type: none"> 1. Interdisciplinarity 2. Inclusiveness 3. Integrity 4. Coherence 	<ol style="list-style-type: none"> 1. Engage with researchers and stakeholders from across disciplinary, geographical, and institutional boundaries. 2. Consider the creation of an intergovernmental panel on AI modelled on the Intergovernmental Panel on Climate Change (IPCC). 3. Redouble efforts towards multilateral cooperation on the regulation of lethal autonomous weapons systems (LAWS). 4. Review internal policies, regulations, and governance arrangements within the UN system with AI in mind.

1. The complex nature of AI regulation: common dilemmas

Any attempt to regulate the development, adoption, use and evaluation of AI technologies requires in-depth consideration, negotiation, and decision-making across multiple dimensions. We highlight **five dimensions that we think pose crucial questions** for AI regulation. This account is intended neither to be exhaustive nor to offer conclusive solutions. The aim is, rather, to identify and delineate the major tensions and sources of contestation in regulatory policy. The levels of complexity and disagreement characterising these debates, in turn, underlines the need for **interdisciplinary perspectives and methodologies**³, as well as opportunities for **fuller democratic oversight and deliberation at the global level**.

Dilemma 1: Definitions

Artificial intelligence is a contested concept. Policy instruments, legislative proposals and technical standards employ varied terminology. Definitions of AI tend to relate to **(1) the types of outputs produced (e.g., predictions), (2) the techniques employed (e.g., machine learning) and (3) the level of autonomy exercised**. Each of these components is complex and gives rise to possible contestation. As a result, competing definitions of AI often reflect the interests of different actors and groups.

AI often takes the form of **dynamic, adaptive, and integrated systems**, not static products with linear processes, clear actors, and discrete outputs.⁴ Foundation models, in particular, have the potential to develop well beyond the intentions of their initial developers.⁵ Though many discussions of the effects of using AI focus attention on Machine Learning (ML), several software techniques are often used in combination in highly complex value chains. Moreover, there are many examples of highly consequential automated decision-making systems that use no ML techniques.⁶ As for autonomy, rather than being an inherent property of a system, it is better understood as a contextual and relational factor. That is, the level of autonomy exercised by a system will **largely depend on how it is embedded and configured** within a given social and institutional setting.⁷

These complexities, and the various interests of different actors, make the task of defining AI for the purpose of regulation both **technically challenging** and **politically charged**. This points to the need for highly interdisciplinary approaches to evidence gathering, decision-making, implementation, and monitoring.

³ Hendrickx, V. and Smuha, N., 2023, [Artificial Intelligence and interdisciplinarity: an evaluation](#).

⁴ Edwards, L., 2022, [Regulating AI in Europe: four problems and four solutions](#).

⁵ Küspert, S., Moës, N. and Dunlop, C., 2023, [The value chain of general-purpose AI](#).

⁶ Rachovitsa, A. and Johann, N., 2022, [The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch SyRI Case](#).

⁷ Beckers, A. and Teubner, G., 2022, [Three Liability Regimes for Artificial Intelligence](#).

Box 2: Interdisciplinarity

AI as a field of inquiry and practice is inherently interdisciplinary. The work of theorising, designing, building, operating, applying, and critiquing AI technologies inevitably involves multiple perspectives, skills, and forms of expertise. Many of the risks AI poses can only be studied and understood with insights drawn from different bodies of knowledge. A computer scientist, for example, may not have knowledge and understanding of complex social inequalities. A sociologist, conversely, may not have the knowledge about advanced machine learning techniques needed to identify underlying problems in the design of AI systems. Working together, they may learn from each other and even develop new framings and approaches to their work.

As we have found in our own research and teaching, this is no easy task.⁸ It requires us to go beyond an additive approach whereby disciplines are effectively considered alongside each other. Instead, we must develop new and hybrid vocabularies, methodologies, concepts, and experts⁹ to produce research that amounts to more than the sum of its parts.¹⁰

We understand interdisciplinarity as cross-disciplinary, cross-sectoral, and cross-cultural; as work that encompasses and integrates academic disciplines as well as public policy, practice, and the expertise of affected communities.

Dilemma 2: Purpose

The proposed European Union (EU) AI Act is perhaps the best-known example of a general regulatory instrument on AI. It creates a tiered regulatory regime with requirements varying according to the level of perceived risk associated with different AI systems. **States and other public authorities elsewhere need not follow the path taken by the EU,** which is bound by its own unique constitutional arrangements and corresponding political goals – namely the promotion of a European single market.¹¹

Alternative models for AI regulation may assume entirely different approaches, and bring within their scope different products, activities, and entities. Policymakers should be **clear about their own objectives and priorities,** and how they can be met by their chosen regulatory model. Many scholars and civil society groups have, for example, highlighted **the importance of international human rights law as the basis for effective regulation**

⁸ MacLeod, M., 2018, [What Makes Interdisciplinarity Difficult? Some Consequences of Domain Specificity in Interdisciplinary Practice.](#)

⁹ Hoffmann, S., Deutsch, L., Klein, J.T. and O'Rourke, M., 2022, [Integrate the integrators! A call for establishing academic careers for integration experts.](#)

¹⁰ Lyall, C., Bruce, A., Tait, J., Meagher, L., 2011, [Interdisciplinary research journeys: practical strategies for capturing creativity.](#)

¹¹ Mazur, J., and Wloch, R., 2023, [Embedding digital economy: Fictitious triple movement in the European Union's Artificial Intelligence Act.](#)

of AI.¹² Although the EU's proposal aims to promote the protection of human rights, alternative regulatory models developed in future could more comprehensively advance the rights of citizens, including social and economic rights.¹³ Global investors have recently called for stronger human rights protection in AI regulation. Such protection would include human rights due diligence, in line with the UN Guiding Principles for Business and Human Rights.¹⁴ Human rights considerations are of **particular importance in migration governance.**¹⁵ This is a field in which the UN has a potentially significant role to play in promoting international cooperation, and in norm-setting.¹⁶

Digital transformation entails profound changes to constitutional orders.¹⁷ Private actors, such as gatekeeper online platforms, govern spaces and relationships that remain formally private while, at the same time, taking on functions traditionally carried out by public authorities. In the public sector, AI systems are often procured from, and developed by, private companies. Sound regulatory policy must be informed by interdisciplinary knowledge of these **shifting public-private dynamics.**

Dilemma 3: Sectoral shape

A key debate is unfolding in AI regulatory policy around how best to address different sectoral needs. The **broadly horizontal (sector-neutral)** shape of the proposed EU regulation can be contrasted with a **more vertical (sector-specific) model**, such as China's regulatory regime.¹⁸ The main advantages of a horizontal approach that applies to all AI systems across sectors are **uniformity, effective coordination, and greater certainty** for citizens who interact with AI. It also avoids the risk of regulated actors either seeking to minimise their obligations by "regulator shopping" or being unsure about which rules to apply.¹⁹ However, critics would contend that it is less flexible than a vertical model that allows policymakers to tailor regulations to different use cases and types of AI. In practice, few regulatory regimes will be entirely vertical or entirely horizontal.

¹² McGregor, L., Murray, D., and Ng, V., 2019, [International Human Rights Law as a Framework for Algorithmic Accountability](#).

¹³ Human Rights Watch, 2021, [How the EU's Flawed Artificial Intelligence Regulation Endangers the Social Safety Net: Questions and Answers](#).

¹⁴ Investor Alliance for Human Rights, 2023, [Citing the Significant Human Rights Risks Inherent in AI, Investors Offer Recommendations to Strengthen the EU's Proposed AI Act](#); United Nations, 2011, [Guiding Principles on Business and Human Rights](#)

¹⁵ Molnar, P. 2022, [The EU's AI Act and its Human Rights Impacts on People Crossing Borders](#).

¹⁶ Fournier-Tombs, E., 2021, [Towards a United Nations Internal Regulation for Artificial Intelligence](#); Barnett, M. and Finnemore, M., 2004, [Rules for the World International Organizations in Global Politics](#).

¹⁷ Micklitz, H-W., Pollicino, O., Reichman, A., Simoncini, A., Sartor, G., and De Gregorio, G. 2022, [Constitutional Challenges in the Algorithmic Society](#).

¹⁸ O'Shaughnessy, M. and Sheehan, M., 2023, [Lessons From the World's Two Experiments in AI Governance](#).

¹⁹ Ada Lovelace Institute, 2021, [Regulate to innovate A route to regulation that reflects the ambition of the UK AI Strategy](#).

Box 3: The healthcare sector

The healthcare sector has been hailed as one area in which AI offers great promise. If successful, AI systems present opportunities to support the work of clinicians, improve medical diagnosis and treatments, and ultimately help produce better outcomes for patients. In most countries, medicines and medical devices are already heavily regulated relative to other objects of regulation. The adoption of AI in health and medicine, however, comes with many clinical, ethical, and other risks. These include risks of errors which result in harm to patients, biases which reinforce existing health inequalities, and security vulnerabilities.²⁰ One such example is the presence of machine bias or discrimination based on legally protected characteristics such as sex/gender or ethnicity, which has often been identified as a contributing factor to inaccuracy and inequalities.²¹ Nevertheless, in some instances, data on these characteristics could serve as a critical component in understanding diseases and enabling improved diagnosis and treatments.

In the UK, a report published in 2022 by the Regulatory Horizons Council found that there is "an urgent need – and an exciting opportunity - to get the regulation right for AI as a Medical Device (AIaMD)."²² Given the unique ethical and technical challenges in healthcare, this may require a more sector-specific and targeted regulatory intervention.

²⁰ European Parliament, 2022, [Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts](#).

²¹ Cirillo, D., Catuara-Solarz, S., Morey, C. et al., 2020, [Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare](#).

²² Regulatory Horizons Council, 2022, [The Regulation of Artificial Intelligence as a Medical Device](#).

Box 4: The public sector

The adoption of AI within the public sector presents specific considerations. Public authorities are generally bound by different obligations than private actors and, in most jurisdictions, must act according to legal and democratic norms. This might include requirements for transparency in aspects of decision making, provision of redress mechanisms for citizens, and restrictions on certain activities such as surveillance.

There is growing evidence that the use of AI in public-service delivery can cause serious harm and often lacks safeguards and mechanisms for accountability.²³ Indeed, many of the most controversial applications of AI have been in public service delivery – for example in social benefits and in policing. Governments also face challenges around limited resources and capacity to develop or procure AI systems that can operate in the public interest.

The Council of Europe’s Committee on Artificial Intelligence (CAI) has published its draft Convention on Artificial Intelligence²⁴ that primarily addresses the responsibilities of public authorities. If adopted, this Convention will be open for accession to member states of the Council of Europe and to non-member states around the world.

Dilemma 4: Prescriptiveness

One of the first things any government must decide when developing a regulatory regime is its **degree of prescriptiveness**. This may range from classical command and control regulation, underpinned by legal penalties and potential criminal liability, to non-regulatory approaches, such as good-practice guidance or reputational incentives. In between are forms of voluntary self-regulation, enforced self-regulation and economic regulation such as awarding of licences.²⁵ Different approaches may also be used in combination, and some approaches may be more **principles-based, rather than using detailed rules**. Here, lessons can be drawn from scholarship and practice in other regulatory spheres, such as financial services, and from historical examples.²⁶

Choices made about prescriptiveness will largely depend on the purpose and objectives of regulatory interventions. They will also have wider policy implications, particularly for **institutional design**²⁷ and **the operation of redress mechanisms**. This will in turn entail policy options about whether, and how, people impacted by AI systems can challenge outcomes and access remedies.²⁸ It also engages various considerations about resource and capacity constraints.

²³ Ada Lovelace Institute, AI Now Institute and Open Government Partnership, 2021, [Algorithmic Accountability for the Public Sector](#).

²⁴ Council of Europe Committee on AI, 2022, [Revised Zero Draft \(Framework\) Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law](#).

²⁵ National Audit Office (UK), 2021, [Good practice guidance Principles of effective regulation](#).

²⁶ Black, J., 2008, [Forms and Paradoxes of Principles Based Regulation](#).

²⁷ Stix, C., 2022, [Foundations for the future: institution building for the purpose of artificial intelligence governance](#).

²⁸ Kaminski, M. E. and Urban, J. U., 2021, [The Right to Contest AI](#).

Dilemma 5: Extra-regulatory tools

In addition to mandatory requirements set out in legislation, policymakers can meet their objectives by promoting the use of a range of tools, including assurance techniques and technical standards. In a less prescriptive regulatory regime, such tools are likely to play an important role.

The term AI assurance is often used to refer to the range of services and mechanisms for checking and verifying AI systems against criteria set out in regulation, standards, and normative frameworks. These include, for example, **audits, impact assessments and conformity assessments**. Legislation may stipulate or incentivise specific assurance activities.²⁹

Policymakers should be aware of the limitations of different assurance techniques, including their methodological limitations. Although audits may be increasingly popular, interdisciplinary scholarship has cast doubt on their efficacy for achieving accountability. Particular attention ought to be paid to the suitability of AI auditors, and the potential for biases by AI auditors themselves (see Box 5).

A technical standard is a document “established by a consensus of subject matter experts and approved by a recognized body that provides guidance on the design, use or performance of materials, products, processes, services, systems, or persons.”³⁰ Policymakers and regulatory authorities may choose to explicitly reference technical standards as a means through which regulatees can demonstrate conformity with the essential requirements of a regulation. The proposed EU AI Act, for example, effectively **delegates much rule-making power to standard-setting bodies**. The degree of this reliance on bodies governed by private law is far from uncontroversial. Although standards are, in theory, intended to address technical specifications, critics have highlighted that even nominally technical safety standards entail value-laden choices about thresholds of acceptable risk.³¹ Problems of **participation, representation, and informational and power asymmetries** persist in standards development organisations (SDOs). Given the importance of standards and SDOs³² and their interplay with legislation, these are important considerations for the democratic oversight and legitimacy of any regulatory regime.

²⁹ Mökander, J., Axente, M., Casolari, F. et al., 2021, [Conformity Assessments and Post-market Monitoring: A Guide to the Role of Auditing in the Proposed European AI Regulation](#).

³⁰ ISO, 2023, [Standards in our World](#).

³¹ Veale, M. and Borgesius, F. Z., 2021, [Demystifying the Draft EU Artificial Intelligence Act](#).

³² Graz, J-C., 2019, [The Power of Standards: Hybrid Authority and the Globalisation of Services](#).

Box 5: Auditing AI

The challenge related to the suitability and eligibility of AI auditors stems from the fact that many significant AI ethical issues are **socio-technical issues that require multidisciplinary perspectives**. For example, tackling the problem of machine bias would involve computer science, moral philosophy, law, and psychology, amongst other disciplines. Consequently, experts from different disciplines may have **different or even contradictory opinions** about these issues, even at the definitional level,³³ which could lead to inconsistency in auditing. To address this challenge, we propose that trained and qualified algorithm auditors should have an interdisciplinary background, or that interdisciplinary teams of auditors be employed. It is imperative for their reliability that consistent workflows and comprehensive auditing frameworks are established.

Challenges also arise from AI auditors' cognitive biases. In behavioural economics, human beings are regarded as bounded rational creatures³⁴ and frequently employ heuristics i.e., a set of thinking shortcuts, which occur unconsciously to make decisions in our daily lives. These heuristics often lead to adequate decision making but may also sometimes lead individuals' decision making to deviate from basic logic, mathematical and probabilistic rationality, or norms, i.e., so-called cognitive bias.³⁵ Algorithm auditors are also prone to these cognitive biases. Individual auditors may hold personal beliefs and stereotypes towards certain groups of people or objects, which may lead to **discriminatory or distorted auditing results**.

It is important that AI auditors are **trained and accredited**. One approach is to mandate that they justify their auditing procedure. From a psychological perspective, this compels individual auditors to engage in conscious decision-making, thus reducing the automated cognitive biases that may arise from unconscious decision-making.³⁶ Additionally, a **third-party observer** should review the auditors' decision process to improve objectivity. Several psychological testing and training programmes are available to assist auditors in identifying their implicit biases and improving decision-making. These include the implicit association test,³⁷ and risk literacy programmes that can improve individuals' statistical inference capabilities in a wide range of settings including health, climate, and finance.³⁸ Whilst these interventions and assessments have merits, we recognize their inherent constraints and limitations.³⁹ Without **sound institutional design** and **meaningful third-party involvement**, AI audits alone are unlikely to be effective and reliable.⁴⁰

³³ Landers, R. N., and Behrend, T. S., 2022, [Auditing the AI auditors: A framework for evaluating fairness and bias in high stakes AI predictive models](#).

³⁴ Simon, H. A., 1990, [Bounded Rationality: Utility and Probability](#); Gigerenzer, G., and Selten, R., 2002, [Bounded rationality: The adaptive toolbox](#); Kahneman, D., 2003, [Maps of bounded rationality: Psychology for behavioral economics](#); Thaler, R. H., 2000, [From homo economicus to homo sapiens](#).

³⁵ Tversky, A., and Kahneman, D., 1974, [Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty](#).

³⁶ Kahneman, D., 2011, [Thinking, fast and slow](#).

Box 6: Responsible AI Licences (RAILs)

The current state of licensing for most open-sourced AI models is dominated by a mix of permissive software licences such as Apache 2.0 or MIT and open data licences such as Creative Commons (CC). While these licences are mainly designed to prioritise rapid development and establish guardrails for ownership of the software or data, **users are given minimal usage restrictions**. These include the freedom to use, modify, redistribute, and build over current work **without substantial emphasis on responsible use**. Permissive licences also do not require developers to adhere to any specific ethical guidelines or principles, which can lead to ethical concerns regarding the use of AI.⁴¹

Responsible AI Licences or RAILs⁴² are a special type of **responsible licence** that acts as a contractual agreement between developers and downstream users who want full access to a model. Specifically, these responsible licences contain **explicit behavioural-use clauses**, which, in contrast to permissive software and data licences, are designed to promote responsible AI practices. They do so by **requiring developers to adhere to specific ethical principles and guidelines**. Moreover, any downstream derivations, including redistribution and transformation from AI models licensed with RAIL, must also abide by use restrictions clauses. This approach means that any AI-based technology can be developed and deployed in a way that is more aligned with specific values and principles, including those set out in instruments like the OECD Principles or in legislation.

³⁷ Greenwald, A. G., McGhee, D. E., and Schwartz, J. L., 1998, [Measuring individual differences in implicit cognition: the implicit association test](#).

³⁸ Hertwig, R., 2017, [When to consider boosting: some rules for policy-makers](#).

³⁹ Laii, C. K., Skinner, A. L., Cooley, E., et al., 2016, [Reducing implicit racial preferences: II. Intervention effectiveness across time](#).

⁴⁰ Costanza-Chock, S., Raji, I. D., Buolamwini, J., 2022, [Who Audits the Auditors? Recommendations from a field scan of the algorithmic auditing ecosystem](#); Raji, D. I., Xu, P., Honigsberg, C. and Ho, D., 2022 [Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance](#).

⁴¹ Keller, P., and Bonato, K., 2023, [Growth of Responsible AI Licensing. Analysis of License Use for ML Models Published on !\[\]\(c380885b89e2c5583288bf9456b72567_img.jpg\)](#).

⁴² Contractor, D., McDuff, D., Haines, J. K., Lee, J., Hines, C., Hecht, B., Vincent, N., and Li, H., 2022, [Behavioral Use Licensing for Responsible AI](#).

2. The global nature of AI regulation: the role of the UN

Amid growing geopolitical tensions and economic competition, multilateral agreements on AI may be unlikely in the short term.⁴³ The **emerging global AI governance landscape is fragmented⁴⁴ and dominated by relatively powerful states and large corporations.** Nonetheless, the **UN has an important role to play** in AI governance and cooperation at the global level, including in the maintenance of international peace and security.

Scholars and other commentators have considered the various **modalities that UN level cooperation on AI could take.** Whilst the creation of a new agency with a mandate for AI cooperation is unlikely to be feasible, an alternative proposition for a body **modelled on the Intergovernmental Panel on Climate Change (IPCC)** would have the advantage of giving policymakers robust assessments of the opportunities, implications, and potential risks of AI.⁴⁵ If this model were to be pursued, we recommend that a strongly interdisciplinary approach be adopted, with representation and participation of researchers and experts from all regions.

The Global Digital Compact is an opportunity for the UN to set out a **coherent and ambitious AI policy agenda of its own.** AI technologies are being used in many fields where UN bodies have specific mandates, including in warfare, migration governance, humanitarian assistance and protection, and climate action. The adoption, development and use of AI presents global challenges that **states alone cannot regulate.** UN bodies can also act as **norm-setters in fields such as migration governance** to influence the behaviour of states.

Within the UN system, there is a clear opportunity to regulate the use of AI through **internal policy and governance.** This might include risk management frameworks, procurement policy, or the development of extra-regulatory tools such as **transparency registers.**⁴⁶

⁴³ Guruparan, K. and Zerk, J., 2021, [Influence of soft law grows in international governance.](#)

⁴⁴ Garcia, E. V., 2020, [Multilateralism and Artificial Intelligence: What Role for the United Nations?](#)

⁴⁵ Mialhe, N., 2020, [AI & Global Governance: Why We Need an Intergovernmental Panel for Artificial Intelligence.](#)

⁴⁶ Haataja, M., van de Fliert, L. and Rautio, P., 2020, [Public AI Registers Realising AI transparency and civic participation in government use of AI.](#)

Box 7: Lethal Autonomous Weapons Systems

The Russo-Ukrainian War has reignited a wide-ranging debate about the use of AI-powered lethal autonomous weapon systems (LAWS) in warfare. The **United Nations Convention on Certain Conventional Weapons (CCW)** debated a ban on autonomous weapons at its review meeting in Geneva 2021 but did not reach consensus.

Notwithstanding this impasse, the UN can take steps towards further cooperation on the regulation of LAWS. In January 2023, a resolution⁴⁷ was adopted by the **Parliamentary Assembly of the Council of Europe (PACE)** which highlighted the need for international regulation of LAWS. It recommended the **development of an international regulation** to ensure appropriate human control, maintain human responsibility and the obligation of accountability, and to implement measures to mitigate risks. While a binding instrument remains a longer-term goal, the PACE recommended that interim steps be taken:

“Pending the emergence of the broad consensus needed to draw up such an instrument, **a non-binding instrument should be prepared in the form of a code of conduct.** This instrument, which might be updated on a regular basis, could codify the guiding principles that are already broadly recognised and highlight the good practices adopted by given States Parties to the CCW.”

⁴⁷ Parliamentary Assembly of the Council of Europe, 2023, Resolution 2485 [Emergence of lethal autonomous weapons systems \(LAWS\) and their necessary apprehension through European human rights law](#).

3. Principles and Recommendations

PRINCIPLES

In recent years, many organisations have issued Principles and Recommendations on the development, adoption, and use of AI. We recognise the importance of instruments, such as the **UNESCO Recommendation on the Ethics of Artificial Intelligence**, which establish shared normative foundations for AI policymaking. Principled commitments alone, however, will not guarantee the realisation of policy objectives.⁴⁸ Policymakers urgently need a stronger **interdisciplinary evidence base** to understand and formulate effective responses to the challenges AI presents for our societies.

Based on our experience as interdisciplinary AI researchers and practitioners, we believe that efforts to regulate AI must be informed by **knowledge that spans disciplinary, geographical, and institutional boundaries**. Multi-perspective knowledge production and exchange can in turn support the development of more coherent, legitimate, and actionable regulatory policy.

1. Interdisciplinarity

AI is inherently interdisciplinary. We must draw on a richer ecology of knowledge to develop, practice, and regulate AI. This requires the creation of new vocabularies⁴⁹ and plural methodologies.

2. Inclusiveness

Regulation must be underpinned by inclusive policy processes and research practices. These processes should draw on the insights and leadership of communities likely to be impacted by AI. This must also be supported by a more equitable distribution of research resources, including funding and compute power.

3. Integrity

AI regulatory policy should be informed by research which is independent⁵⁰ and conducted according to the highest ethical standards.⁵¹ Integrity in AI research also includes considerations about labour exploitation,⁵² data governance, and climate.

4. Coherence

⁴⁸ Mittelstadt, B. 2019, [Principles alone cannot guarantee ethical AI](#).

⁴⁹ AI Now, 2021, [A New AI Lexicon](#).

⁵⁰ Ahmed, N., Wahed, M. and Thompson, N. C., 2023, [The growing influence of industry in AI research](#).

⁵¹ UNESCO, 2017, [Recommendation on Science and Scientific Researchers](#).

⁵² Gray, M., and Suri, S., 2019, [Ghost work: How to stop Silicon Valley from building a new global underclass](#); Shefeni, S., 2023, [The invisible labour of Africa in the Digital Revolution](#).

There is a need to strategically coordinate research activities to better synthesise and translate evidence from research into policy.

RECOMMENDATIONS

1. Policymakers should engage with researchers and other stakeholders from across disciplinary, geographical, and institutional boundaries to develop regulatory policy informed by wide-ranging evidence and expertise. This can help address the most **fundamental dilemmas for AI regulation**, including those related to **definitions, purpose, sectoral shape, prescriptiveness**, and the role of **extra-regulatory tools**.
2. The Global Digital Compact should present options for the most appropriate actors, institutions, and modalities to take global cooperation on AI regulation forward. This should include the possibility of creating an **intergovernmental panel modelled on the Intergovernmental Panel on Climate Change (IPCC)**. Any cooperation mechanism would need to be **inclusive and representative** of experts and **communities from across the globe**, strongly **interdisciplinary**, and fully **independent** in its mandate and membership.
3. All stakeholders should use the opportunity of the Global Digital Compact to redouble efforts towards multilateral cooperation on the **regulation of lethal autonomous weapons systems (LAWS)**.
4. **Internal policies, regulations and governance arrangements within the UN system** should be reviewed with AI in mind. The adoption of AI presents multiple challenges for UN operations and oversight mechanisms.⁵³ We recommend the promotion of an **effective, efficient, and coherent approach to AI regulation across the UN**. In practice, this might include changes to Enterprise Risk Management (ERM) policies,⁵⁴ updated procurement procedures, and new activities to be conducted by the Office of Internal Oversight Services (OIOS). The development of extra-regulatory tools to promote accountability, such as a public AI register, should be considered. These measures would build on the **Principles for the Ethical Use of Artificial Intelligence in the United Nations System** agreed in December 2022.

⁵³ International Telecommunications Union (ITU), 2022, [United Nations Activities on Artificial Intelligence \(AI\) 2022](#).

⁵⁴ United Nations Joint Inspection Unit, 2020, [Enterprise risk management: approaches and uses in United Nations system organizations](#).