Developing Artificial Intelligence Sustainably: Toward a Practical Code of Conduct for Disruptive Technologies

By Gordon Myers and Kiril Nejkov

The adoption and diffusion of artificial intelligence and other disruptive technologies will play an important role in market creation and growth. Development finance institutions have a role to play in leveraging their investments to ensure that these technologies sustain both growth and development objectives. To this end, the authors propose adoption of a Technology Code of Conduct as a framework, supported by a set of practical tools for its operationalization, to assist IFC’s clients engaged in technology intensive projects.

Artificial intelligence (AI) and other disruptive technologies, much like electricity and the Internet, are general purpose technologies (GPTs). GPTs matter for development because they contribute to innovation across the economy, resulting in exponentially higher growth outcomes. Because of their broad impact, the extent to which GPTs are adopted by firms and diffused across a market depends on the quality of institutional settings and support.

A critical factor underpinning the institutional environment for disruptive technologies is trust. Consumers and stakeholders must trust that privacy will be respected, that data are used responsibly, that technologies are adopted in a way that is environmentally and socially sustainable, and that in particular, these technologies are adopted in a way that supports inclusion and equity.

Giving firms the practical risk management tools needed to develop trust is pivotal to achieving the real development promise that disruptive technologies like AI can offer (see EM Compass Notes 69 and 71). In the absence of these tools, the public conversation has focused on the very real concern that AI has a unique capacity to create harm and that AI-based innovation has accelerated beyond regulatory understanding and control. In some cases, absence of trust has already translated into market and political concern. Thus, ensuring trust has emerged as a precondition to realizing the social, commercial, and public benefits of implementing AI technologies.

There are ongoing global efforts to develop principles to guide the ethical development and use of AI and other new technologies. However, more practical guidance on managing AI risks and implementing agreed-upon principles is needed at the firm level. As a development finance institution (DFI) with clients across emerging markets, IFC is well placed to convene stakeholders to develop and deploy a Code of Conduct together with a set of practical tools to ensure such granular and practical guidance. IFC can play such role because of its reach to

About the Authors

Gordon Myers, Chief Counsel, Technology and Private Equity, Legal and Compliance Risk, IFC, and Co-Chair, Legal and Policy Community, ITS Innovation Lab, World Bank Group (gmyers@ifc.org).

Kiril Nejkov, Counsel, IFC Compliance Risk, Office of the Chief Compliance Officer, Legal and Compliance Risk, IFC (knejkov@ifc.org).
clients across emerging markets, its understanding of sustainability principles, its experience in translating these principles into practical operational guidance, and its strong history of collaborating with DFIs on frameworks designed to support sustainable, inclusive, and responsible investment. The purpose of this Note is to explain the authors’ work to date in developing a draft Code and tools, outline the next steps in refining these products, and invite interested parties’ feedback.

What is AI?

AI is the science and engineering of making machines intelligent, especially intelligent computer programs.\(^2\) AI is therefore a series of approaches, methods, and technologies that display intelligent behavior by analyzing their environments and taking actions—with some degree of autonomy—to achieve specific targets that can improve the provision of services.\(^1\)

Toward a Technology Code of Conduct

Over the past two years there is a growing number of initiatives supporting the ethical use of AI. For example, the AI Ethics Guidelines Global Inventory, developed by Algorithm Watch,\(^4\) lists over 80 frameworks, while the Principled Artificial Intelligence project at Harvard University developed a map of 32 ethical and rights-based approaches to AI ranging from civil society, government, international organizations, and multi-stakeholder initiatives to the private sector.\(^5\)

Many of these frameworks reflect similar principles. Some of these principles go back to science fiction author Isaac Asimov’s “three laws of robotics,” articulated in the 1940s.\(^6\) Others seem to reflect prima facie duties developed by a relatively small community of moral philosophers in the 1960s, including beneficence, non-maleficence (do no harm), and truth-telling.\(^7\)

The two most important recent developments are the adoption of OECD Principles on Artificial Intelligence\(^8\) and the European Commission (EC) Ethics Guidelines for Trustworthy Artificial Intelligence.\(^9\) The OECD Principles are the first government-endorsed framework on AI policy. These Principles aim to provide high-level guidance on the development of the overall policy environment for AI at a national level, and to facilitate international knowledge sharing and cooperation. The EC Guidelines are more detailed and include some useful tools such as firm-level risk-assessment questionnaires designed to inform the European Union’s overall approach to AI, including its strategy for investing in disruptive technologies.

The OECD Principles and EC Guidelines address five groups of issues:

- **The aspirational big-picture implications** of AI and how its design, development, and use should contribute to inclusive growth and societal and environmental well-being;
- **The implications of AI on humans** and how to embed human-centric values into the technology, most notably human autonomy and fairness;
- **Transparency and explainability** of both the AI systems and their individual outputs;
- **Technical robustness** of AI systems, ensuring their security and safety;
- **Accountability** for AI technology so that, consistent with ordinary public, legal and policy expectations, institutions and individuals involved in the AI systems lifecycle are held responsible for the operation and outcomes of such systems.

The OECD Principles and EC Guidelines have made a helpful contribution in narrowing the discussion from high-level ethical issues (like philosophical concern with the moral rightness of individuals) to more focused ones (the human-centeredness and trustworthiness of AI).

In the authors’ view, these two frameworks provide an opportunity to consider the broader sustainability and inclusion aspects of AI. The immediate and ongoing benefits of sustainable AI investment can be bolstered through more granular guidance for firms in designing and adopting AI technologies.

The Proposed Technology Code of Conduct

IFC has played an important role in working with industry, investors and other stakeholders to develop standard-setting frameworks. For example, IFC’s Performance Standards on Environmental and Social Sustainability,\(^10\) the Corporate Governance Development Framework,\(^11\) and the Operating Principles for Impact Investment\(^12\) have been adopted as baseline standards across many DFIs and private sector investors.

In recent years, IFC clients have demanded more detailed guidance in managing the risks arising from the adoption of disruptive technologies, particularly in markets where clients move faster than regulators. That is why IFC has taken the initiative to canvass good practices and gaps in existing frameworks, and to take into consideration the needs of IFC’s clients, to develop a draft Technology Code of Conduct (“the Code”) and related tools to operationalize the principles of the Code.
The authors believe that the principles contained in the Code reflect the core values implicit in IFC’s Environmental and Social Performance Standards that define IFC clients’ responsibilities for managing their environmental and social risks (see Table 1) and provide a more useful roadmap to trust and operationalization of advanced technologies, including AI, across markets. In summary, the principles in the Code are organized into three tiers, from most conceptual to most operational:

**The three Core Values**—Benefit, Inclusiveness, and Fairness—provide the underlying, absolute priorities and provide a reference point for resolving potential conflicts and inconsistencies in the implementation of the safeguard and building block-level principles.

**The six Safeguards**—Transparency, Informed Consent, Validation, Security, Responsibility, and Accountability—build on the Core Values and provide a basis for developing concrete, practical tools, processes, and systems needed to achieve outcomes consistent with the Core Values.

Finally, the **three Building Blocks**—Governance, Proportionality, and Continuity—provide the overall...
framework and inform, integrate, and establish realistic expectations for the tools and approaches to be developed.

The authors have informally consulted with technology investors, other DFIs, and internal staff, including investment and ESG experts, for reaction. The reviewers noted that the Code strikes an appropriate balance between being both aspirational and operational and is neither overly intrusive nor unworkably abstract. They also felt that the Code would add value, support public trust, and provide a framework for sustainable investment in AI-based innovation.

The authors incorporated several revisions following this initial feedback. For example, the authors have clarified that compliance with applicable law is required in markets where the legal framework is well developed. At the same time, the Code should facilitate technological neutrality and identify areas for self-regulation. Revisions have also been made to ensure that the Code and the tools can be implemented—especially by investors—in a way that does not create disproportionate financial or operational burdens on investee companies, especially early-stage ones. Importantly, the tools must be flexible enough to be customized across use-cases, for example, by industry, company maturity, and product type.

**Operationalizing the Technology Code of Conduct**

For the Technology Code of Conduct to be impactful, it must be underpinned by practical tools. To develop these tools, the authors analyzed 35 of the over 80 existing AI frameworks that are supported by practical tools and mapped these existing tools against the Code principles. Such mapping was used to identify good practices and gaps and develop a draft framework of key tools that client companies can use to implement the Code. Finally, a draft Progression Matrix was developed, identifying expected practices in relation to each principle of the Code for different stages of company maturity: from minimum acceptable practice for emerging companies, to expected practices for later stage companies, to leadership practices for mature companies.

**Analysis of the good practices and gaps in existing tools**

The authors believe that operationalization of the Code requires tools that address both the technical processes, such as privacy and compliance by design, and the business processes, such as risk governance and reporting. The authors have accordingly grouped the tools in these two categories. Obviously, some of the tools fit both categories, and the originators of the tools may not agree with the proposed classification, which is put forward only for easier analysis for the purposes of this Note.

**The 10 technical tools analyzed** (Table 2) generally address practices related to the principles of Fairness, Transparency, and Validation.

As applied to AI, in relation to Fairness, the technical tools help their users determine whether data is complete and properly formatted and whether the datasets are representative of the AI live environment. For example, there are several notable cases where facial recognition algorithms have been trained on datasets containing a larger percentage of white, male faces, leading them to perform poorly on black women. Similarly, some hiring tools have not scored women fairly for technical positions, such as engineers, by comparing CVs to the existing, predominantly male, universe of engineers.

In relation to Transparency, the technical tools assist with the ability of AI systems to explain their decisions in a way that is comprehensible to humans. Examples include describing how lending algorithms take social-media connections into account and give better rates to people with “higher quality” networks.

Lastly, in relation to Validation, the technical tools assist with human interpretation of AI outputs by regularly testing for inaccuracies or discrimination in an AI system’s conclusions and developing plans for responding to user complaints or potential harm caused by the AI system.

The authors’ analysis also suggests that the technical tools analyzed do not yet sufficiently address the principles of Benefit, Responsibility, Accountability, and most notably, Governance. The draft Code and supporting tools focus on filling these gaps.

**The 25 business process tools analyzed** (Table 3) are much more diverse, both in terms of their format and their target users. Given that the business tools are much more comprehensive, the authors have focused on consolidating and leveraging the good practices embodied in these tools. For example, existing tools for policy makers are useful references for developing good practice guidance for operationalization of the Responsibility principle. Similarly, the existing business process tools that target technology professionals can provide guidance on how to comprehensively address end-to-end risks within the overall AI project development cycle—from design and development checklists to auditing tools.

Lastly, there is also room for better customization of some of the well-known business and investment-decision making tools (for example, BCG matrix, Porter’s Five, SWOT, and PEST) to more systematically address the specific risks of AI applications.
Progression Matrix and Model Documents

Bearing in mind the above identified key gaps and good practices of the existing AI tools, the authors propose development of a Progression Matrix and selected model documents to operationalize the Technology Code of Conduct.

The Progression Matrix provides details on the technical and business processes that client companies should adopt, in relation to each of the principles of the Code and subject to client companies’ financing stage and maturity. An extract of the Matrix is available at the end of this Note (Annex 1) and the full Matrix is available here: www.ifc.org/EMCompassNote80A_TCoC-Matrix. The Progression Matrix details the practical implications of the Code’s principles for different types of companies and anticipates development of good practice model guidance for adopting companies. Some of the model documents the authors plan to develop include:

- **Well-Being Impact Assessment**—a conceptual framework (for example, based on Maslow’s hierarchy of needs) to identify beneficial and unbeneﬁcial components of AI systems.\(^\text{13}\)

- **Stakeholder Engagement Plan**—with defined principles of stakeholder engagement, stakeholder mapping, and practical steps.\(^\text{14}\)

- **Bias Impact Assessment**—a set of questionnaires for assessment, review, and disclosure of AI systems’ potential impact on fairness among affected communities.\(^\text{15}\)

- **Privacy by Design**—a reference framework for privacy protection and information management that can be applied from specific technologies to whole information ecosystems and governance models.\(^\text{16}\)

Conclusion and Next Steps

Adoption of the Technology Code of Conduct will help build trust with customers and ensure that AI technologies are human-centric. The Code will also support sustainability and impact, including investment by the impact investing community, by ensuring that AI systems contribute to the well-being of individuals, society, and the environment.

---

**TABLE 2 Selected Technical Tools to IFC Technology Code of Conduct—Public Draft**

*Source: IFC.*

<table>
<thead>
<tr>
<th>Toolkit Format</th>
<th>Target Users</th>
<th>Relevance to Code of Conduct</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Aequitas Bias and Fairness Audit Toolkit</td>
<td>▲ ✔ ✔</td>
</tr>
<tr>
<td>2</td>
<td>AI Explainability 360 Open Source Toolkit by IBM</td>
<td>▲ ✔ ✔</td>
</tr>
<tr>
<td>3</td>
<td>AI Fairness 360 Open Source Toolkit by IBM</td>
<td>▲ ✔ ✔</td>
</tr>
<tr>
<td>4</td>
<td>Deon</td>
<td>▲</td>
</tr>
<tr>
<td>5</td>
<td>Fairness Flow by Facebook</td>
<td>▲</td>
</tr>
<tr>
<td>6</td>
<td>Fairness Tool—Accenture</td>
<td>▲</td>
</tr>
<tr>
<td>7</td>
<td>LF AI Foundation ML workflow</td>
<td>▲</td>
</tr>
<tr>
<td>8</td>
<td>Lime</td>
<td>▲</td>
</tr>
<tr>
<td>9</td>
<td>TransAlgo</td>
<td>▲</td>
</tr>
<tr>
<td>10</td>
<td>What-If Tool—Google</td>
<td>▲</td>
</tr>
</tbody>
</table>
The authors invite feedback on the proposed Code and Progression Matrix to ensure that the Code reaches its full potential as a practical tool for sustainable innovation. This is to incorporate and balance, from the outset, input from the investor, company, regulator, and civil society communities, before engaging in more formal and systematic consultation. The next step would be to test the draft principles and guidance, in a sandbox-type partnership with selected client companies, to confirm the practical usefulness and impact of our approach. The authors’ hope is that donors, impact investors, and other DFIs will see value in cooperating in these efforts.

Finally, the authors hope to launch the Code iteratively, to assure rapid rollout and continuous improvement by supporting establishment of a community of stakeholders, similar to the Equator Principles.17 Expressions of interest in participating in these efforts is welcome. Please send all feedback and expressions of interest to ifctechnologycode@ifc.org

---

**TABLE 3 Selected Business Process Tools Mapped to IFC Technological Code of Conduct—Public Draft**

Source: IFC.
ACKNOWLEDGMENTS

The authors would like to thank Guangji Xue, Compliance Analyst, IFC Compliance Risk, Office of the Chief Compliance Officer, IFC, for his support in the research, and the following colleagues for their review and suggestions: John Salmon, Partner, Hogan Lovells International LLP, London, UK; Atiyah Curmally, Principal Environmental Specialist, Environmental, Social & Governance—Global Advisory, Advice and Solution, Global Industries, IFC; Charles Travis Canfield, Principal, Corporate Governance, Advice and Solutions, Global Industries, IFC; Piotr A. Mazurkiewicz, Principal Environmental Specialist, Environmental & Social Policy—Risk Management, IFC; Monique Mrazek, Senior Investment Officer, Manufacturing, Agribusiness and Services, Latin America and the Caribbean Region, Health & Education, IFC; Felicity O’Neill, Global Partnerships Associate, Partnerships and Multilateral Engagement, IFC; within Thought Leadership, Economics and Private Sector Development, IFC: Baloko Makala, Consultant; Research Assistants Prajakta Diwan and Kevin Matthees; and Thomas Rehermann, Senior Economist.

Please see the following additional reports and EM Compass Notes about technology and its role in emerging markets and private investments in infrastructure:

Reinventing Business Through Disruptive Technologies—Sector Trends and Investment Opportunities for Firms in Emerging Markets (March 2019); Blockchain: Opportunities for Private Enterprises in Emerging Markets (January 2019); Artificial Intelligence and the Future for Smart Homes (Note 78, forthcoming); Artificial Intelligence and 5G Mobile Technology Can Drive Investment Opportunities in Emerging Markets (Note 76, Dec 2019); How Artificial Intelligence is Making Transport Safer, Cleaner, More Reliable and Efficient in Emerging Markets (Note 75, November 2019); Bridging the Trust Gap: Blockchain’s Potential to Restore Trust in Artificial Intelligence in Support of New Business Models (Note 74, Oct 2019); Artificial Intelligence: Investment Trends and Selected Industry Uses (Note 71, Sept 2019); The Role of Artificial Intelligence in Supporting Development in Emerging Markets (Note 69, July 2019). Blockchain and Associated Legal Issues for Emerging Markets (Note 63, Jan 2019).

6 Asimov wrote the thee Laws of Robotics for a short science-fiction story which was later included in his book I, Robot published in 1950: First Law: A robot may not injure a human being or, through inaction, allow a human being to come to harm; Second Law: A robot must obey the orders given it by human beings except where such orders would conflict with the First Law; Third Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.
7 Sir William David Ross, in his book The Right and the Good, listed seven prima facie duties: fidelity, reparation, gratitude, justice, beneficence, non-maleficence, and self-improvement.
8 https://www.oecd.org/going-digital/ai/principles/
10 https://www.ifc.org/wps/wcm/connect/topics_ext_content/ifc_external_corporate_site/sustainability-at-ifc/company-resources/sustainable-finance/equator+principles+financial+institutions
11 https://www.ifc.org/wps/wcm/connect/topics_ext_content/ifc_external_corporate_site/ifc+cg/cg+development+framework
12 https://www.impactprinciples.org/
13 See for example: IEEE Well-being Impact Assessment Concept. Maslow’s hierarchy of needs is a theory in psychology which defines a five-tier model of human needs from physiological to self-actualization.
17 IFC’s Performance Standards on Environmental and Social Sustainability, which define IFC clients’ responsibilities for managing their environmental and social risks, have emerged as globally recognized standards for management of environmental and social risks. They serve as a basis for the Equator Principles, launched in 2003, and are currently adopted by more than 90 banks and financial institutions. https://www.ifc.org/wps/wcm/connect/topics_ext_content/ifc_external_corporate_site/sustainability-at-ifc/company-resources/sustainable-finance/equator+principles+financial+institutions

This publication may be reused for noncommercial purposes if the source is cited as IFC, a member of the World Bank Group.

The complete Progression Matrix can be found as EM Compass Note 80A here: Note 80A

### Core Values

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. BENEFIT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technology should provide customers, individuals, and communities with access to products, services, and capabilities that benefit them.</td>
<td>Commerciably viable product providing benefits to as many customers, individuals, and communities as possible</td>
<td>Same</td>
</tr>
<tr>
<td></td>
<td>No inherent harm that cannot be sufficiently minimized, mitigated, or responsibly accepted in the context of the industry and relevant social norms</td>
<td>Clearly articulated purpose of using the technology for the benefit of individuals, communities, and the environment; and not causing harm to individuals, communities, or the environment. All potential benefits and risks, with relevant mitigants, clearly documented</td>
</tr>
<tr>
<td></td>
<td>Product performs intended function consistently and correctly</td>
<td>Product design addresses potential indirect and negative longer-term impact of adoption</td>
</tr>
<tr>
<td></td>
<td>Same, and guards against immediate negative side effects of technology incorporated into product design</td>
<td>Product continually updated to maximize benefit based on user experience and feedback</td>
</tr>
<tr>
<td></td>
<td>Product continually updated to maximize benefit based on user experience and feedback</td>
<td>Full Well-Being Impact Assessment (for example, based on the Maslow Hierarchy of Needs) developed, with regular updates</td>
</tr>
</tbody>
</table>

### Additional EM Compass Notes Previously Published by IFC Thought Leadership

**FEBRUARY 2020**

**Note 79:** Accelerating Digital Connectivity Through Infrastructure Sharing

**Note 78:** Artificial Intelligence and the Future for Smart Homes

**JANUARY 2020**

**Note 77:** Creating Domestic Capital Markets in Developing Countries: Perspectives from Market Participants

**DECEMBER 2019**

**Note 76:** Artificial Intelligence and 5G Mobile Technology Can Drive Investment Opportunities in Emerging Markets

**NOVEMBER 2019**

**Note 75:** How Artificial Intelligence is Making Transport Safer, Cleaner, More Reliable and Efficient in Emerging Markets

**OCTOBER 2019**

**Note 74:** Bridging the Trust Gap: Blockchain’s Potential to Restore Trust in Artificial Intelligence in Support of New Business Models

**Note 73:** Closing the SDG Financing Gap—Trends and Data

**SEPTEMBER 2019**

**Note 72:** Blended Concessional Finance: The Rise of Returnable Capital Contributions

**Note 71:** Artificial Intelligence: Investment Trends and Selected Industry Uses

**AUGUST 2019**

**Note 70:** How Insurtech Can Close the Protection Gap in Emerging Markets