Witzel, Mardi; Bhargava, Niraj

Working Paper

AI-related risk: The merits of an ESG-based approach to oversight

CIGI Papers, No. 279

Provided in Cooperation with:
Centre for International Governance Innovation (CIGI), Waterloo, Ontario

Suggested Citation: Witzel, Mardi; Bhargava, Niraj (2023) : AI-related risk: The merits of an ESG-based approach to oversight, CIGI Papers, No. 279, Centre for International Governance Innovation (CIGI), Waterloo (Ontario)

This Version is available at:
https://hdl.handle.net/10419/299976

Centre for International
Governance Innovation

CIGI Papers No. 279 — August 2023

AI-Related Risk

The Merits of an ESG-Based Approach to Oversight

Mardi Witzel and Niraj Bhargava

Centre for International
Governance Innovation

CIGI Papers No. 279 — August 2023

AI-Related Risk
The Merits of an ESG-Based Approach to Oversight

Mardi Witzel and Niraj Bhargava

About CIGI

The Centre for International Governance Innovation (CIGI) is an independent, non-partisan think tank whose peer-reviewed research and trusted analysis influence policy makers to innovate. Our global network of multidisciplinary researchers and strategic partnerships provide policy solutions for the digital era with one goal: to improve people's lives everywhere. Headquartered in Waterloo, Canada, CIGI has received support from the Government of Canada, the Government of Ontario and founder Jim Balsillie.

À propos du CIGI

Le Centre pour l'innovation dans la gouvernance internationale (CIGI) est un groupe de réflexion indépendant et non partisan dont les recherches évaluées par des pairs et les analyses fiables incitent les décideurs à innover. Grâce à son réseau mondial de chercheurs pluridisciplinaires et de partenariats stratégiques, le CIGI offre des solutions politiques adaptées à l'ère numérique dans le seul but d'améliorer la vie des gens du monde entier. Le CIGI, dont le siège se trouve à Waterloo, au Canada, bénéficie du soutien du gouvernement du Canada, du gouvernement de l'Ontario et de son fondateur, Jim Balsillie.

Credits

Managing Director of Digital Economy Robert Fay
Director, Program Management Dianna English
Project Manager Jenny Thiel
Senior Publications Editor Jennifer Goyder
Publications Editor Susan Bubak
Graphic Designer Brooklynn Schwartz

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org

Table of Contents

About the Authors

Mardi Witzel is an associate with NuEnergy.ai and is focused on ESG (environmental, social and corporate governance) and AI governance, and the special challenges facing high-growth firms. She serves on the board of PolyML, a private firm specializing in machine learning and advanced analytics, and has 20 years of experience in not-for-profit board governance, stakeholder engagement and strategic planning.

Niraj Bhargava is the CEO and lead faculty at NuEnergy.ai and an expert on artificial intelligence (AI) governance. He leads a team of experts specialized in AI governance education, in the creation of organization-level AI governance frameworks and in the integration of NuEnergy's AI trust measurement software, the Machine Trust Platform.

Acronyms and Abbreviations

AGI	artificial general intelligence
AI	artificial intelligence
AIA	Algorithmic Impact Assessment
AIDA	Artificial Intelligence and Data Act
AI RMF	AI Risk Management Framework
CDSB	Climate Disclosure Standards Board
ESG	environmental, social, governance
IFRS	International Financial Reporting Standards
IIRC	International Integrated Reporting Council
IP	intellectual property
ISO	International Standards Organization
ISSB	International Sustainability Standards Board
IT	information technology
NGOs	non-governmental organizations
NIST	National Institute of Standards and Technology
SASB	Sustainability Accounting Standards Board
TCFD	Task Force on Climate-Related Financial Disclosure
VRF	Value Reporting Foundation

Executive Summary

Artificial intelligence (AI) technology has been found to generate value for many firms; however, it also has unintended and undesirable consequences. The reality of AI-related risk has led to the development of AI governance frameworks and calls for greater oversight of the use of AI.

The merits of an ESG (environmental, social, governance)-based approach to oversight of AI-related risk are considered in this paper, with a focus on the current trajectory of international sustainability standards development. Despite their differences, AI governance and ESG reporting both seek to address risk in the broadest sense, with proactive and transparent approaches to its management and mitigation.

Recognizing that readers may be familiar with either AI governance or ESG but not likely both, the paper is constructed so as to provide an overview of each. The paper examines what is different about AI-related risk and identifies four factors: speed and scale, AI empowerment, AI life cycle and AI ethics. The analysis finds possible gaps and/or material topics that are not covered by the Sustainability Accounting Standards Board (SASB), including AI-related risks that may differ on an industry basis and on the basis of an enterprise's role in the AI value chain. A preliminary set of recommendations for incorporating material AI-related risk into ESG reporting, covering both general or context-setting disclosures and industry-specific disclosures, is provided.

Introduction

AI technology is being applied broadly in business. According to one study (McKinsey & Company 2022), adoption has more than doubled since 2017, but the proportion of organizations using AI has plateaued in recent years at 50–60 percent. Another global survey (IBM 2022) revealed a continuing upward trend in adoption, with 35 percent of companies reportedly using AI today, 42 percent exploring its use and AI adoption up four percent from 2021 to 2022. And while AI has been found to generate value for many firms, it is also recognized as a double-edged sword, giving rise to a host of unintended and undesirable consequences.

Reflecting the reality of AI-related risk, there has been a proliferation in the development of AI governance frameworks and a growing body of literature calling for greater oversight of its use. A recent article in *MIT Sloan Management Review* (Silverman 2020) identified several lenses through which boards may approach the management of AI-related risk: compliance, strategic planning, legal or business risk, and ESG.

This paper examines the merits of an ESG-based approach to board oversight of AI-related risk. This is a practical and timely discussion. Following a vanguard of early AI adoption in the late 2000s, the hard reality of the technology's risks and negative impacts became apparent. In response, regulatory authorities took notice, and various legal and policy instruments have been proposed, passed or are under review. This takes nothing away from the enormity of the potential for AI to do good, to promote growth through innovation and increased productivity, and, with any luck, to help humans address some of the biggest challenges our planet and society face. But it is fair to say that alongside these great expectations, is a now equally rooted belief in the imperative of responsible or trustworthy AI.

In recent years, there have been developments in both AI governance and sustainability reporting that share common themes. The emphasis on the twin objectives of leveraging opportunity while managing risk is evident from both, albeit to differing degrees. The focus of AI governance is more pronounced on risk, while the driving force behind sustainability reporting is a bit of a dance between the two. Most critically, AI governance and sustainability reporting have both advocated strongly for greater transparency and disclosure across a broader set of risks. Each discipline, if they can be labelled as such, is concerned with the identification, management and mitigation of risks that extend beyond short-term financial impacts, including legal, regulatory and reputational concerns, for example. In particular, ESG has pushed the agenda of evaluating opportunity and risk according to a longer-term view. Where historically these types of risks may have been seen as non-financial, the proponents of ESG hold that these should be seen as pre-financial risks that are destined to come home

to roost in enterprise valuation and financials at some point — and often sooner than later.

The discussion relating to AI as an ESG risk has the potential to become complex, and this paper has intentionally set some boundaries. The first is an acknowledgement that while AI is part of the broader constellation of digital governance concerns, and many of the arguments it contains could be applied to data and digital more broadly, this paper remains focused on AI. The second boundary pertains to the literature on ESG and sustainability: these reference both single materiality (i.e., the issues that pose material financial risks to a firm) and double materiality (i.e., the consequences of a firm's operations that pose material impacts to the environment and society); this paper is focused only on a discussion of the financial materiality of AI risk to firms.

Despite their different histories, the worlds of AI governance and ESG reporting share a common motivation: to address risk in the broadest sense, with proactive and transparent approaches to its management. Evidence of this agenda can be seen in recent developments from each discipline. The publication of the European Union's proposed Artificial Intelligence Act (AI Act) in April 2021, represents the first example of a comprehensive regulatory approach to AI, carrying with it a broad suite of obligations, including transparency and disclosure relating to AI, and, in particular, the enterprise systems supporting its responsible development and use. In November 2021, the International Financial Reporting Standards (IFRS) Foundation announced the creation of an International Sustainability Standards Board (ISSB), signalling the advent of a more unified global approach to voluntary reporting and disclosure on ESG or sustainability standards.[1]

The root of the case for an ESG-based approach to assessing risk and opportunity is found in stakeholder orientation, the time horizon it applies to both risk and opportunity and the role of reporting and disclosure. Stakeholder concern about AI-related business impacts, in general, and how data is sourced, secured and used, in particular, is at an all-time high. The call for appropriate corporate management and disclosure of AI use is emerging as both a public expectation and, in many jurisdictions, a legal or regulatory requirement. Whereas conventional accounting methods are limited today in terms of their ability to incorporate many financially material issues, ESG frameworks provide a complementary approach.

The literature on ESG and AI governance has been evolving, but there are only a few papers specifically exploring the utility of ESG as a framework for understanding, reporting and disclosing AI-related risk. James Brusseau (2023) finds the current suite of ESG ratings frameworks lacking for evaluating AI impacts and proposes a model based on commonly held principles for ethical AI, rather than adaptation of an existing ESG framework. Henrik Skaug Sætra (2021) proposes a framework for evaluating ESG-related impacts of AI according to the United Nations Sustainable Development Goals. Sætra (2022) builds on earlier work, presenting an AI ESG protocol — a framework for evaluating the ESG implications of AI capabilities, assets and activities according to three scopes of impacts and where these are experienced in the supply chain.

This paper builds on this research in an examination of the value of applying an ESG lens to the challenge of AI governance, and specifically the management of AI-related risk. Recognizing the practical implications of the newly formed ISSB for global reporting, the contribution of this paper is to review the range of AI governance tools available today, explore the nature of AI-related risk and set forth an approach to how the reporting and disclosure of AI-related risk could be integrated into the work of the ISSB, toward development of a single global baseline for sustainability reporting. The hope and expectation is a fully implemented ESG framework, incorporating AI and, ultimately, the full realm of digital governance, will result in systems, controls and accountability for monitoring and reporting on the part of chief financial officers.

The first section provides an overview of what is meant by ESG, because any evaluation of its value as a lens requires a basic appreciation of what ESG is. With this in hand, the second section conceptually explores the question of why AI should be governed through an ESG approach. In order to dig into this question more deeply, the third and fourth sections examine the state of AI governance approaches and the nature of AI-related risk, respectively. Having a foundational understanding of AI-related risk and AI governance facilitates the discussion that is the meat of this

1 The terms ESG and sustainability are used interchangeably in this paper in relation to the reporting and disclosure initiatives.

research, found in the last two sections of the paper. The fifth section addresses the idea of materiality in the context of AI-related risk and the sixth section provides preliminary ideas about how to accommodate these material concerns into the structure of ESG standards, reporting and disclosure that exists today. Specifically, the paper proposes questions and ideas reflecting the state of the ISSB's guidance as captured in its Exposure Draft, General Requirements for Disclosure of Sustainability-related Financial Information.[2]

What Is ESG?

It is hard to find a good single definition of ESG, because it depends on the application. ESG might be seen as a set of environmental, social and governance criteria that investors use to screen investments. It could be seen to represent the score of a firm's collective consciousness for ESG factors. Alternatively, a company might view ESG as a set of standards for corporate behaviour to be used in formulating strategies for long-term value creation.

ESG is a framework for thinking more comprehensively and, therefore, more accurately about the risks and opportunities that firms face over short, medium and longer time horizons and how these may impact firm performance. Regardless of what it is called, there is a strong likelihood that important risks and opportunities are not being managed if a firm has not adopted an ESG framework. Instead of calling it ESG, this approach to thinking, planning, doing and reporting could instead be called "integrated financial and non-financial governance" because it incorporates a broader range of factors impacting company performance and valuation than traditional rules of financial disclosure require today.

Practically speaking, ESG for organizations manifests itself as a management, reporting and disclosure approach that may be facilitated by one or more of a number of ESG frameworks. Large public companies are especially likely to be doing ESG reporting today, with evidence that more than 90 percent of the S&P 500 publish sustainability reports (Governance and Accountability Institute 2021), although the transparency and quality of their data vary considerably. It is not only public companies — private companies, government agencies and non-governmental organizations (NGOs) also publish ESG data. Organizations may choose to use one of the leading global frameworks, such as those provided by the Global Reporting Initiative or the Value Reporting Foundation (VRF)[3] to identify, measure and report on the issues that are most material to their businesses. And beyond the information provided directly by an organization, investors and other stakeholders may look to sustainability information that is published by third-party ESG Ratings Agencies, such as Sustainalytics or MCSI ESG Research.

The number of companies that publish ESG reports will continue to grow, as will the depth and quality of data, for a few reasons. Stakeholders, including investors, consumers, employees and regulators, are demanding access to this information. Additionally, there is evidence of ESG's value as a driver of firm performance and of the role intangible assets play in enterprise valuation (often eclipsing tangible assets). And while the current hodgepodge of global ESG standards and frameworks has undermined uptake across many sectors, that is about to change, with the announcement of a global initiative to create universal standards for sustainability or ESG reporting. In 2021, the IFRS Foundation established an ISSB as a parallel organization to the International Accounting Standards Board, whose financial reporting standards are used in over 140 countries. This new organization will spearhead convergence on a set of harmonized global sustainability standards, in collaboration with the world's leading ESG reporting frameworks. The new ISSB will consolidate leading investor-focused sustainability disclosure organizations including the Climate Disclosure Standards Board (CDSB) and the VRF, which itself is an amalgam of the former SASB and the IIRC. This consolidation, into a body capable of developing and overseeing a single set of global sustainability standards, is expected to be a game changer for sustainability reporting.

If a company is taking a proactive approach to ESG, it means the board and senior management team are, one, consciously asking questions, getting

2 See www.ifrs.org/projects/work-plan/general-sustainability-related-disclosures/exposure-draft-and-comment-letters/.

3 The Sustainability Accounting Standards Board (SASB) and the International Integrated Reporting Council (IIRC) came together to form the VRF in 2021.

educated and building plans that incorporate ESG factors alongside traditional product-market mix planning; and, two, making capital expenditures in both areas. In the environmental, or "E," space, that might involve planning around things such as climate resilience, energy management, waste reduction or ecological impacts. In the social, or "S" space, it is the organization's practices and impacts in relation to human and social capital that are considered — human rights, health and safety, diversity and inclusion, customer privacy, and data security are examples. Finally, in the governance, or "G," space, the issues at the forefront include business model and innovation, business ethics, management of the legislative and regulatory environment, and systemic risk management.

What is an ESG factor that a company may not be thinking, planning or doing anything about today but where it matters? Take planning for net-zero. In the E space, there is all kinds of talk about the UN 2030 Agenda for Sustainable Development and the push for net-zero emissions by 2050, but has this translated into concrete plans within most firms? Or in the S space, how about an organization-wide data strategy, focused on the collection, security, application and governance of data? And if a firm does in fact have a well-constructed plan for ESG-type factors, with deliverables, timeframes and aligned systems such as compensation policy, is there a value to that? Why would a robust net-zero plan providing competitive differentiation and growth potential not be worth something — a net positive contributor to firm value?

Despite the controversy and the skeptics, it seems inevitable that the global energy sector will go through a major transition away from carbon-intensive products and toward more sustainable solutions, representing a major economic opportunity. On the flipside, if an organization is not practising good data governance, is it then not only failing to capitalize on an opportunity, but also exposing itself to future risk? It seems logical that firms that are early to the party with well-conceived plans and capital expenditure strategies will be beneficiaries, and laggards who focus only on traditional financial analysis will fall behind. Concrete planning approaches in areas such as climate resilience and data governance can be seen as intangible assets, in the same way a piece of intellectual property (IP) is.

The point of all this is to highlight the materiality of ESG factors and their relationship to both the non-financial and financial performance of firms. Across the spectrum of intangible asset types, AI could be viewed as a sort of "levered" driver of firm valuation, helping firms make the best use of their other tangible and intangible assets. It has been argued that AI is upending the industrial age and destroying traditional business thinking (Davenport 2019). New commercial behemoths are growing up without the legacy anchors of inflexible physical assets, and many of the most valuable companies today are software, network and platform-based, and have little in the way of physical assets. For both new and traditional business models, AI represents a powerful tool for value creation.

Why Govern AI Use through an ESG Approach?

Why approach the governance of AI through ESG? To answer this question, it helps to understand what is behind the momentum of ESG today, and specifically the role of ESG in supporting long-term value creation. ESG or sustainability approaches have emerged at the frontier of corporate purpose and strategy, enabling both value creation and risk mitigation for firms. The rise in ESG is as much explained by the changing composition of market valuations as by some sense of moral imperative. Taking stock of a company's ESG scorecard maps closely to its non-financial performance, and that is a very big deal today. Why? Because the non-financial elements of firms are growing in both size and proportion, and exerting major influence on firm valuation.

There has been a dramatic shift in the proportion of corporate valuation that is attributable to intangible versus tangible assets. Assets such as human capital, IP, company reputation and customer loyalty represent greater value than they used to. According to one recent study, only 17 percent of the value of S&P 500 market value was attributed to intangibles in 1975, whereas by 2015 that number had grown to 84 percent (Ocean Tomo 2021). The materiality of intangible assets is reinforced both by the scale of their economic

presence and empirical evidence of their positive correlation with higher price-to-book ratios.

What does this have to do with AI? AI is increasingly seen as an asset — a driver of organizational value. AI may be employed as a lever to help firms operate more efficiently and effectively and, in this sense, it may be viewed as a productivity tool, just another technology for information technology (IT) to manage, but it is more than that. AI is a pervasive technology, potentially with applications in every department, entangled with data and enterprise system architecture. AI is capable not just of doing things faster and better than humans, but of doing things on a scale that previously would have been impractical for humans. It is already and will continue to be a game changer in many industries. In this sense, AI has the potential to be among any firm's most valuable intangible assets.

At the same time, AI is not without risk. There are risks relating directly to the implementation of AI, including the possibility of data mismanagement, algorithmic bias, error and drift. There are risks associated with the complexity of AI and the challenge of explaining outcomes. There are first-order effects from AI use that may impact individuals or organizations. People may be discriminated against or wrongly directed. Companies and industries may experience job losses. There are also second-order consequences associated with AI use, including dynamics such as skills atrophy, with the potential to impact individuals, organizations and society.

In practical terms, the ESG elevation of broadly based reporting and disclosure practices serves to inform investors, instill discipline and shape capital investment decisions in consideration of both near-term and longer-term material factors. The reality is that regulation and standards never entirely keep up with innovation, and there is a requisite for good firm-level governance to reach beyond the sufficiency of compliance requirements. But it is also true that regulation and standards can provide meaningful guidance and goalposts to organizations. And today, with the launch of the ISSB and the promise of a forthcoming set of global standards for sustainability reporting, the practice of identifying, measuring, managing and reporting on a wide range of financial and non-financial information is likely to become institutionalized.

The bottom line is AI use represents both an opportunity and a risk — its use is a material concern for firms, with the potential to influence financial performance, firm reputation and to impact stakeholders. The widespread recognition of AI as a double-edged sword has spawned an industry around the call for its responsible use. These initiatives may be referred to as "responsible AI" or "ethical AI" and while they acknowledge the enormity of the technology's business potential, the focus is generally more around how AI is governed from a risk management standpoint.

The World of AI Governance

AI governance is an overarching term that is used to refer to a wide range of approaches that have emerged in response to the perceived risks and impacts associated with AI. The potential for AI-related risks and impacts is broadly acknowledged today, with the policy and practices around it continuing to evolve at the international, national and subnational levels of government and among myriad industry and NGO actors. This section provides an overview of some of the highlights in AI governance developments — strategies, policy, guidance, regulations, standards and practices — relating to the recognition of AI-related risk.

National Strategies

The Organisation for Economic Co-operation and Development catalogues hundreds of national AI policy initiatives from 69 countries, territories and the European Union.[4] These are broadly categorized into four groups: governance-related (564), financial support (294), AI enablers and other incentives (423) and guidance and regulation (301). These initiatives are being undertaken by government entities, research and education organizations, private companies, social groups, individual economic actors (for example, entrepreneurs and private investors) and intermediaries (for example, incubators, industry associations and technology transfer offices).

4 See https://oecd.ai/en/dashboards/overview.

Ethical AI Frameworks

Numerous organizations have proposed ethical principles for the development and deployment of AI technologies. These include well-known frameworks such as the European Union's Ethics Guidelines for Trustworthy AI, the Institute of Electrical and Electronics Engineers' Global Initiative for Ethical Considerations in AI and Autonomous Systems, the Partnership on AI's Principles for AI and the Singapore Model AI Framework. These principles-based frameworks are each unique but they tend to share common threads. Most of the leading ethical AI frameworks include principles relating to transparency and explainability, fairness and bias, accountability, privacy and human-centricity.

Legal and Regulatory Instruments

Until recently, legislation relating to the development and use of AI was limited to subnational, sector-specific cases. That changed with the introduction of several overarching AI laws proposed at the national level. AI-related legislation will evolve over time given the dynamic nature of the technology itself. If there was any question of this, the inevitability was laid bare with the rise of generative AI as a popular tool, and the consequences for how AI should be defined. The players in the value chain and who should be held accountable for different controls and responsibilities are important to consider.

- **EU Proposed AI Act:** Proposed in April 2021, the EU AI Act represents the first example of a comprehensive regulatory instrument for AI oversight at the national — or, in this case, supranational — level of government. The EU AI Act takes a risk-based approach to regulating AI, outlining four different risk categories: unacceptable risk, high-risk, limited risk and minimal risk. The bulk of the proposed regulation addresses requirements for high-risk systems, which include robust approaches to risk management, data governance, technical documentation, record-keeping, transparency and provision of information to users, human oversight and accuracy, robustness and cybersecurity. The burden of responsibility is mainly placed on the developers of AI, who have an obligation to implement a quality management system and other stipulations, according to the guidance of the proposed regulation. This graduated tiering of requirements based on risk is often referred to as "proportionality," marking efforts by regulators to focus the regulatory burden where it is most needed and free less risky application spaces to pursue innovation unfettered. Numerous standards are contemplated by the European Union to support their legislative objectives. A recent standardization request from the European Union to CEN-CENELEC (the European Committee for Standardization and the European Committee for Electrotechnical Standardization) proposes standards in 10 areas covering AI, organizational systems, data quality and data access.

- **US Algorithmic Accountability Act:** In 2022, US lawmakers introduced the Algorithmic Accountability Act in both the House and Senate. The Algorithmic Accountability Act is focused on automated processes and systems deployed to render "critical decisions." Within two years of enactment, the proposed act will require the Federal Trade Commission to promote regulations that require impact assessments. Based on the outcomes of these impact assessments, covered entities will be required to undertake actions to eliminate or mitigate impacts that demonstrate a material negative impact that is expected to have a legal or other significant effect on a consumer's life.

- **Canada's Artificial Intelligence and Data Act (AIDA):** In 2022, the Canadian federal government introduced Bill C-27 in the House of Commons for first reading. As a part of Canada's Digital Charter, Bill C-27 contains three separate acts relating to data privacy and AI, including the proposed AIDA. AIDA is focused on the provision of new rules for the responsible development and use of AI. A key component of the bill outlines that companies must assess whether their AI systems are "high impact" (to be fully defined in the regulations) and, if so, they must meet a set of obligations around risk assessment and mitigation of bias, system monitoring, transparency and record keeping, notice and the use of anonymized data. The proposed AIDA is a component of Canada's National AI Strategy, which was launched in 2017 and updated in 2020. The strategy aims to promote the development and adoption of AI in Canada while also addressing issues such as ethical considerations, diversity and inclusion, and the impact of AI on the workforce.

- **Canada's Algorithmic Impact Assessment (AIA):** The AIA is a mandatory risk assessment tool for use within the federal government, in support of the Treasury Board's Directive on Automated Decision-Making. An impact assessment section is at the heart of the tool, querying matters such as the level of human involvement in the directive, reversibility of a decision and duration of impact. The output of the AIA is the determination of the impact level of an automated decision-making application as Level I (little to no impact), Level II (moderate impact), Level III (high impact) or Level IV (very high impact). Based on the impact level, there are different requirements for peer review, notice, human-in-the-loop for decisions, explanation requirement, training, contingency planning and approval for the system to operate.

AI Standards

Numerous standards organizations are developing standards in support of the responsible use of AI. Below are examples from two international leaders, the International Standards Organization (ISO) and the National Institute of Standards and Technology (NIST).

- **ISO 42001 — AI Management System:** ISO 42001 provides requirements for establishing, implementing, maintaining and continually improving an AI management system. Organizations are expected to focus application of requirements on features that are unique to AI. ISO 42001 outlines the need for organizations to understand their internal and external context and the needs of interested parties, and to establish the scope of the management system on this basis. The standard sets out requirements in six different areas:
  - leadership (for example, AI policy, roles, responsibilities and authorities);
  - planning (for example, risk criteria and risk assessment, system impact assessment, AI objectives);
  - support (for example, resources, competence, awareness, communication, documentation);
  - operation (for example, processes for operational planning and control);
  - performance evaluation (for example, monitoring, measurement, evaluation, internal audit); and
  - improvement (for example, continual improvement of AI management system).
- **NIST AI Risk Management Framework (AI RMF):** The AI RMF acknowledges the unique risks associated with AI systems and promotes risk management as a key component in the responsible development and use of AI. According to NIST, AI risk management can drive responsible uses and practices by prompting organizations and their internal teams that design, develop and deploy AI to think more critically about context and potential or unexpected negative and positive impacts. The NIST AI RMF includes four functions to help organizations address the risks of AI systems in practice: govern, map, measure and manage. While "govern" applies to all stages of organizations' AI risk management processes and procedures, the "map, measure and manage" functions can be applied in AI-system-specific contexts and at specific stages of the AI life cycle. The NIST AI RMF characterizes the principles for trustworthy AI as accurate, valid and reliable, safe, explainable and interpretable, privacy-enhanced, fair, and accountable and transparent.

  NIST has also issued a special publication titled *Towards a Standard for Identifying and Managing Bias in AI*, in which it discusses three categories of bias (systemic bias, human bias, and statistical and computational bias) and provides updated life cycle guidance (pre-design, design, deployment).

Enterprise-Level AI Governance

The review above highlights the significant level of activity in the guidance side of the AI governance space today. A different question is what are organizations actually *doing* in regard to AI opportunity and risk? Anecdotal evidence suggests that the proliferation of principles-based frameworks and guidance has seeped into the corporate ethos, with many companies, organizations and government departments espousing their own set of ethical AI principles. Where the challenge now lies is in the operationalization of this guidance. Knowing what ethical AI *looks* like is a very different matter to knowing *how* to do it, much less

actually implementing the processes, procedures and accountability regimes to achieve it.

AI governance leaders recognize the reality that one size does not fit all. Best-in-class approaches to enterprise-level governance take stock of the array of AI governance guidance and involve the creation of organization-specific frameworks for ethical AI and the development of plans and practices for operationalizing these frameworks. The call to operationalize AI governance principles and frameworks represents a significant challenge for enterprise AI.

The challenge of operationalizing ethical AI principles with practical approaches is at the core of enterprise-level AI governance. Moving from "principles to practice" is a common refrain in AI governance today, and a fundamental weakness of the array of regulatory approaches is the fact they have not stimulated much in the way of practical implementation. In many respects, the world of ESG and sustainability reporting is going through a similar evolution, but a little further ahead. The work of the Task Force on Climate-Related Financial Disclosure (TCFD) for example, represents a determined intent to construct a framework that would facilitate implementation as well as disclosure. The TCFD recommendations are not narrowly focused on an organization's emissions metrics, but include disclosure requirements relating to governance, strategy, risk management and targets. The ISSB's draft guidance is building on these same disclosure pillars and is expected to propagate more uptake and better governance of sustainability topics.

The progress that is being made on enhanced sustainability reporting should be helpful to the cause of AI governance on two fronts, operationalization and standardization. The direction of ESG reporting provides an example of how to nudge things along the spectrum from principles-based frameworks to guidance that incorporates expressions of accountability, operational requirements and metrics. In addition, ESG reporting is on the verge of having something else to offer instructionally, through the standardization of sustainability frameworks across global jurisdictions. The drive for global benchmarking of sustainability-related risks and opportunities to help investors make informed decisions has highlighted the merits of institutionalizing an approach that both satisfies investors' information needs and promotes the discipline of good governance.

The "global sustainability standards train" is leaving the station and there will soon be global benchmarks and standards for a wide range of sustainability-related risks and opportunities. It seems imprudent that AI, with its promise of massive economic and social impact going forward, should be left out. AI governance can learn something from the progression of sustainability reporting and disclosure and advance these practices operationally. For this to happen, there needs to be a robust discussion around the nature and materiality of AI risk and opportunity and the implications for reporting and disclosure. The next section focuses on the nature of AI-related risk.

The Nature of AI-Related Risk

Contemporary dictionaries define AI as a subfield within computer science where the practice is one of machines developing the capabilities of humans, including cognition, pattern recognition, reasoning and decision making, and performing tasks in a human-like way. The terms AI, machine learning, deep learning and neural networks are often used interchangeably, but the processes by which they operate are different and create different levels of risk for organizations. For many companies, it is a subset of AI known as machine learning that holds the most interest and where the bulk of AI development is taking place. Natural language processing, neural networks and deep learning are all subsets of machine learning. Generative AI is also a form of machine learning.

The array of tools and techniques that comprise machine learning are often referred to as artificial narrow intelligence. Beyond machine learning and deep learning is artificial general intelligence (AGI). AGI is a bigger concept that involves the creation of intelligent machines that can simulate human thought and behaviour. Little progress has been made developing higher forms of AI that would approach decisions much in the same way humans do, understanding or remembering emotions and interacting with people. Even the advent of widely available generative AI applications, such as the

Open AI ChatGPT application, are examples of narrow rather than general AI. ChatGPT does not think or feel like a human — it is trained to scan and synthesize information from an unfathomably vast online database, producing collated material from existing sources, more like a smart digital librarian than a philosopher or inventor.

In this paper, the terms machine learning and AI are used interchangeably, as is colloquial in society today. In all cases, they refer to artificial narrow intelligence. Business managers do not need to understand the technical details of AI, but they do need to be armed with enough knowledge to know what is happening in their organizations. This includes knowing what type of AI is being used and for what purposes, what risk each application carries and what approach the firm is taking to manage that risk. This should be a proactive exercise and should start with an understanding of the nature of AI-related risk.

What Is Different about AI-Related Risk from a Technology Standpoint?

Organizations have looked to technology for decades to support their business operations. Financial services, in particular, have used models to reduce risk, support regulatory compliance and realize strategic business needs. Traditional models, for example, have been used in banking to support decisions and predictions in areas such as capital provisioning, strategic planning, pricing, asset liquidity, customer relationship management, money laundering and fraud detection. What is it about AI that amplifies the risk scenario from traditional decision models?

Machine-learning models have delivered a host of benefits through their ability to scale up decision making or predictions and address tasks that are beyond traditional models due to the amount of data involved and the constraints of time. Somewhat ironically, the availability of big data is one of the factors that has facilitated growth in machine learning, along with access to greater computing power and progress in algorithmic development. But with these developments come trade-offs. In contrast to traditional models, machine-learning models are dynamic and non-deterministic. Their capacity to learn is the foundation of their ability to operate at scale, but this also introduces variability into the process. Machine-learning models can go off-track.

In machine learning, the model is typically designed to describe, predict or prescribe something — these are known as target variables or outcomes. How do machines learn to do this? Machine learning starts with data, and it may include spreadsheets, text, pictures, financial transactions, readings from sensors or the information in annual reports. The bulk of the data available is used to train the machine-learning model and a small portion of it is reserved to test the performance of the model's accuracy. The risks and unintended consequences that emerge from the core machine-learning application itself derive from three areas: the handling of data, characteristics of the model and deployment of the model over time.

- **Data-related risk:** Leveraging data for AI is a business-driven call, about harnessing the latent power in data to get at strategically valuable information. This creates a need to make data broadly available across the enterprise, and introduces distinct risks in terms of data storage, data security and data privacy. Additionally, the machine-learning process has its own challenges, with data potentially migrating across multiple partners, organizations and/or countries. The presence of different laws, policies and ethics at each stage may impact model learning. Something that is considered ethical in one jurisdiction may be considered unethical in another.

- **Characteristics of the machine-learning model:** The risk deriving from the machine-learning model may relate to model quality, model complexity or the approach to model learning and validation. Model quality is compromised by the misapplication of machine-learning algorithms to find patterns in data where none exist. Model complexity increases as design passes from machine learning to deep learning. The flip side of the power of neural networks is their opaqueness, generating outcomes that may not be easily explainable. Model-learning style, and specifically the use of supervised versus unsupervised learning, underlies the inherent risk in machine learning. With unsupervised learning there is no outcome variable on which to train the model, and the algorithm is on its own to sort through data for patterns and structures. This creates a different management and oversight challenge.

- **Life cycle of the model and model drift:** One of the sources of AI-related risk derives from the AI life cycle — the progression of steps that includes data activities, model design, model development, model deployment, monitoring for model drift and model disposal. There are different risks with each stage of the life cycle and the inevitability of data drift presents a continuous challenge. Machine-learning models are dynamic and this is what really distinguishes machine-learning risk. Data changes because the world changes. A model trained on winter seasonal temperatures will make different predictions than one trained on summer temperatures. Because of drift, some models need retraining, even on a regular basis. In machine learning, model drift is normal; it is something to be expected and managed, but it provides a moving target.

The risks associated with machine-learning technology have implications for its use in operational settings and this demands new governance approaches. Machine-learning risks today are real, but the potential for really bad things to happen is more related to management and oversight, than any mystical capacity of the technology itself. There is nothing super-human about today's AI technology that should enable it to escape human control and direction. The generative AI applications present a trickier challenge because of the vastness of the data on which the models are trained and on which they run and continue to learn. But even here, in theory, human developers have control of what data the models are fed and the ability, with labelling, to parse out undesirable data elements. Machine learning is a tool, and like other powerful tools throughout human history, it can be used for good or bad, and even without intention, its use can have negative consequences. This underlies the call for trustworthy AI.

What Is Different about AI-Related Risk from an Organizational Standpoint?

In addition to the technology-related risks of AI, there are risks and implications to its development and use in an organizational context, as a tool for practical application. There are four factors characterizing the risks associated with AI in the context of its practical use:

- **Speed and scale:** The speed and scale with which AI can drive outputs might be viewed as what underlies the opportunity for this technology, but also presents an exacerbating condition for risk. AI done well yields enormous upside. Conversely, AI done badly can be very bad for business — not to mention people and the planet. This reality is the motivating force both to embrace the opportunity of AI and cultivate robust AI governance.

- **AI as empowering:** AI is not a passive technology. It harnesses the latent power of data toward some end that will have an effect. Placing this tool into the pipeline of an entity's operation means there is going to be an effect on something — an individual, a group, the planet, the company, a stakeholder. The fact AI is, by definition, evidence based, is likely to empower those who have access to it, and promote more automated decision making, for good or bad.

- **The AI life cycle:** The life cycle associated with AI (i.e., from data through design, development, deployment and ongoing monitoring for drift) constitutes a new challenge for governance and oversight. The fact that risks vary at different stages of the life cycle is one risk but the bigger challenge is the inevitability of drift and the persistence of change after deployment. This has implications for the types of policies, processes and organizing structures that firms need to have in place to effectively govern AI.

- **Ethical AI:** There are unique challenges in answering the call for AI that is human-centric, fair and non-discriminatory, transparent and explainable, high quality and accurate, safe and secure, compliant with privacy rules and subject to clear accountability. This has implications for policy makers inside organizations, but also for employees more broadly, as the role of data and data analytics is likely to permeate all aspects of enterprise activity.

The important thing to recognize about AI-related risk is the fact that it derives both from unique aspects of the technology and from the practical realities of how AI is developed, used and managed by individuals and organizations. Alongside the opportunity side of AI, these risks may present and evolve in varying combinations, giving rise to potential legal, regulatory, reputational and financial impacts.

The Materiality of AI-Related Risk

In February 2023, Google shares dropped by nine percent in a single day — a stunning US$100 billion loss in market cap — in the wake of revelations that Bard, its recently released chatbot, had produced a factual error. This was not the first time AI had gone badly with a cost, but it was the first time the world saw a giant stumble and fall hard with AI. There are huge expectations for generative AI and AI in general. The capital market's response to Bard's shortcomings is the literal embodiment of the materiality of AI-related risk.

The idea of materiality is at the core of ESG and sustainability reporting. The materiality of AI-related risk (and opportunity) underlies the argument for incorporating AI into sustainability and ESG frameworks. In its early work, the ISSB has aligned its description of materiality with IFRS Accounting Standards, stating "the IFRS Foundation's focus is on meeting the information needs of investors. Therefore, the ISSB uses the same definition of 'material' that is used in IFRS Accounting Standards — that is, information is material if omitting, obscuring or misstating it could be reasonably expected to influence investor decisions."[5]

The Google share price slide provides concrete evidence of the financial impact of AI-related risk, but it is just one isolated example (AI lacks accuracy), with one particular application (ChatGPT/generative AI). Considering the broader ecosystem of AI technologies and use cases, what can we say constitutes material information? Building on the path the ISSB is carving toward a global baseline of sustainability-related reporting and disclosure, the question that logically follows is: What is it about the AI that firms are developing, procuring, deploying and using, that could reasonably be expected to influence investors' decisions? And, more specifically with regard to AI-related risk: What AI-related risk is material in a sustainability context?

Accepting that there are novel risks associated with AI in terms of both the technology and its practical application, is it reasonable to suggest that the very presence of AI in an organization's value chain is material? More likely, there will be some threshold above which information pertaining to AI-related risk will be material but, critically, this information will include insight into both the technology and the organizational strategy, policies and processes for addressing the risk. And because an organization's ambitions and strategies for AI are typically expressed at the enterprise level, while the technology is most often applied to solve a particular business problem or opportunity, there will be material information at both levels.

The authors suggest there are three different types of disclosure information that may be considered material:

i) enterprise-level information about systemic strategy, policies, processes and procedures;
ii) use case-specific information about policies, processes and procedures; and
iii) measures and metrics of performance for performance evaluation of i) and ii).

Disclosure on Policies, Processes and Procedures (Enterprise Level and Use-Case Level)

While an organization's grand strategy for AI and policies around its development, procurement and use will (hopefully) live at the enterprise level, the reality is many AI projects take place within an entity's business units, where AI is applied to solve specific business problems. The AI use cases that interest business units may be totally different, with different technology and governance implications. This means there will be information that is material to reporting and disclosure at both the enterprise-level and the use-case level. How an entity is organized in terms of accountability, roles and responsibilities and according to what policies, processes and procedures is material to the topic of AI-related risk.

Performance Measures and Metrics

Information about the performance of the enterprise writ large, and pertaining to its specific AI use-case systems is material. For each level, there should be associated measures and metrics for evaluation of performance as compared to benchmarks and goals.

5 See www.ifrs.org/groups/international-sustainability-standards-board/issb-frequently-asked-questions/.

These metrics will need to capture information about the performance of the technology in the traditional sense of model accuracy, but also across a range of measures for trustworthy AI, including topics such as transparency, explainability, fairness and bias. Information about both an organization's goal aspirations and its performance versus goals will be material to enterprise value and, therefore, to investors.

Materiality of AI-Related Risk and Direction of the ISSB

The idea of scoping the materiality of AI-related risk according to organizational systems and processes, on the one hand, and performance metrics, on the other, aligns with the ISSB's guidance in its Exposure Draft (S1) "General Requirements for Disclosure of Sustainability-related Financial Information."[6] The ISSB's draft guidance describes sustainability-related financial information as broader than the information reported in financial statements, and potentially including an entity's governance of sustainability-related risks and opportunities: the strategy for addressing them, the expected impact of related decisions on cash inflows and outflows, the entity's reputation, performance and prospects as a consequence of related actions and its development of knowledge-based assets.

The ISSB's draft guidance identifies four areas of core disclosure content: governance, strategy, risk management, and metrics and targets as they each relate to an entity's approach to identifying, addressing, managing and monitoring significant sustainability-related risks and opportunities. The Exposure Draft (S1) represents the ISSB's overarching set of general draft guidance, and it is understood that more specific standards will follow, addressing discrete sustainability topics and, possibly, addressing these on an industry-basis as appropriate. In the meantime, the ISSB's guidance instructs entities to use the draft guidance and "consider the disclosure topics in the industry-based SASB Standards, the ISSB's non-mandatory guidance (such as the CDSB Framework application guidance for water- and biodiversity-related disclosures), the most recent pronouncements of other standard-setting bodies whose requirements are designed to meet the needs of users of general purpose financial reporting, and sustainability-related risks and opportunities identified by entities that operate in the same industries or geographies" (IFRS 2022).

SASB Standards in an AI Context

Materiality is about preparing businesses to address risks and opportunities, and it follows that the issues of greatest impact in mining precious metals are not the same as those in health care or banking. Likewise, the risks for AI in law enforcement, the judicial system and health care are not the same as in entertainment or manufacturing. For this reason, it is likely that an industry-specific approach to AI-related sustainability standards will be appropriate to complement the ISSB's four areas of core disclosure content. The SASB framework lends itself to industry- and issue-specific disclosures, as individual standards are outlined across 77 industries in six sectors, and are characterized according to five dimensions: environment, social capital, human capital, business model and innovation, and leadership and governance.

For companies working with AI, there are potential impacts in each SASB dimension. This raises two obvious questions:

- Do the SASB standards capture disclosure requirements for anything material to AI that is not covered by the ISSB's draft guidance?
- Do the SASB standards, as they exist today, adequately cover the sustainability-related risks and opportunities associated with AI?

It is difficult to answer the first question without a clearer sense of how organizations will use the new ISSB draft guidance, but, on balance, the character of the ISSB disclosure requirements looks different to SASB. The disclosure requirements in the draft guidance read like requests for information about macro, enterprise-level systems and processes for oversight. In contrast, SASB queries more specific topics such as emissions, human rights impact and consumer privacy, and does this on an industry-by-industry basis. There will be some overlap for sure, especially on questions relating to business model impact and governance topics, which make an appearance in

6 The ISSB's Exposure Draft (S1) "General Requirements for Disclosure of Sustainability-related Financial Information" was released in March 2022 for public comment. It is designed to be the sustainability equivalent of IAS 1 "Presentation of Financial Statements" that defines a complete set of financial statements and IAS 8 "Accounting Policies, Changes in Accounting Estimates and Errors" that provides guidance on the establishment and implementation of accounting policies.

both frameworks. Nonetheless, there are reasonable grounds to believe SASB will function as a nice complement to the ISSG general requirements.

Turning to the second question, then, are the SASB standards of today adequate? This question can be broken down into several sub-questions, each of which will need to be addressed individually:

a) Are there material topics relating to AI risk (and opportunity) that are not currently captured by the SASB framework?
b) Is the materiality of AI-related risk (and opportunity) different industry-to-industry?
c) Is there a divide between the material issues facing inherently digital AI companies and/or use cases (for example, fraud detection) versus AI production applications (for example, advanced robotics in manufacturing)?
d) Is there a difference in the materiality of risks (and opportunities) for the developers versus procurers versus users versus platform providers of AI?

The first three questions, captured in points a), b) and c), are really asking the question should there be a new (horizontal) topic(s) in the SASB framework pertaining to AI? The final question, posed in point d), is more nuanced: If it is decided that AI developers face different material risks, does that point to a discrete set of questions in a(nother) new horizontal topic or, alternatively, are AI developers sufficiently different to other software companies, such that a new vertical "AI developer industry" might be warranted?

Point a) — Sufficiency of Current SASB Standards for AI

The authors contend that the SASB standards do not adequately capture AI-related risk and opportunity today. The SASB requirements for software and IT services best illustrate this because, arguably, this is the industry with the most overlap with AI. The SASB standards ask firms in the software and IT services industry to measure and report on six metrics in four issue categories: environmental footprint of hardware infrastructure (environment); customer privacy and data security (social capital); employee engagement, diversity and inclusion (human capital); and competitive behaviour and system risk management (leadership and governance).

Let us focus on the customer privacy and data security, again, offering lots of overlap with AI. Figure 1 shows that the SASB standards for software and IT services include a set of quantitative metrics and a request for description of policies and practices, for both customer privacy and data security.

Currently, the request is for policies, practices and performance on key metrics relating to data privacy and data security *in a generic way*. In the context of AI, and assuming applications where the level of risk or impact crosses the threshold to warrant reporting, there would be different questions. These questions would pertain to data, models and outcomes, *specifically in the context of an AI use case*, and potentially incorporating queries relating to the different stages of the AI life cycle. These might include disclosure requirements pertaining to the trustworthiness of AI, including explainability, fairness, bias and accuracy.

Additionally, there would be questions to extract information about an entity's enterprise-level systems for governing AI including quality management, risk management and privacy management approaches.

These types of questions and measures, as they apply to AI, are absent from the current SASB framework. Further analysis is likely to uncover numerous topics for consideration across multiple dimensions. AI is a pervasive technology and can be expected to have impacts on individuals, society, the planet, business models and governance. These impacts will be both positive and negative — there will be opportunities and risks. Once the ISSB decides to evaluate AI-related risks and impacts, new and expanded challenges for framing both risk and opportunity are likely to be uncovered.

Points b) and c) — Differences in Materiality of AI by Industry

It seems intuitive that the risks and opportunities associated with AI will vary, but an interesting question is whether they vary by industry or use case, and how SASB (or any other framework) would handle that. Superficially, it might appear that a bank would have material issues relating to AI whereas a manufacturing company would not. But what if the manufacturing company is not just using AI in robotics, but is also using it to screen resumes, or for a program relating to physical safety on the plant floor? And how quick should we be to

Figure 1: Software and IT Services Sustainability Disclosure Topics and Accounting Metrics, Data Privacy and Freedom of Expression, Data Security

Data Privacy & Freedom of Expression	Description of policies and practices relating to behavioral advertising and user privacy	Discussion/Analysis
	Number of users whose information is used for secondary purposes	Quantitative
	Total amount of monetary losses as a result of legal proceedings associated with user privacy	Quantitative
	(1) Number of law enforcement requests for user information, (2) number of users whose information was requested, (3) percentage resulting in disclosure	Quantitative
	List of countries where core products or services are subject to government-required monitoring, blocking, content filtering, or censoring	Discussion/Analysis
Data Security	(1) Number of data breaches, (2) percentage involving personally identifiable information (PII), (3) number of users affected	Quantitative
	Description of approach to identifying and addressing data security risks, including use of third-party cybersecurity standards	Discussion/Analysis

Source: SASB (2018).

dismiss the materiality of AI in robotics? The high probability is that within the next decade, most industries and every company of a certain size will be using AI and facing associated material risk and opportunity. This reinforces the requirement for AI-related topics as horizontal additions to the framework, and suggests the variability with which the framework may apply them across industries, which SASB is well-suited to accommodate.

Point d) — Differences in Materiality of AI by Type of AI Actor

For starters, let us make it simple and consider only whether the materiality of risks and opportunities may be different for developers versus users of AI. This question might point to the merits of horizontal topic additions to the SASB framework, to accommodate the particular risks and opportunities that AI developers face. Alternatively, there could be an argument that AI development, at least for those organizations that are intensely involved in pioneering AI development, warrants its own new vertical. This would be the case if this hypothetical vertical of intense AI developers was found to have a sufficiently unique set of material disclosure topics that other industries — principally software and IT, but also generally any industry that is using AI and doing little in-house development — do not.

This question warrants further investigative analysis and is, in the fullest sense, beyond the remit of this paper; however, we offer preliminary thoughts. Appreciating that the SASB framework considers materiality according to industry sector and capital dimension, there could be an analysis of the distinct risks that the use of AI development poses within each of the five dimensions and how these are distinct to those of AI deployers and users. The table below provides a sample of the types of questions that stakeholders, including regulators, investors and members of the general public, may want insight into.

So where does all this leave us? We have wide-ranging guidance from the world of AI governance where there is no standardization but common themes around principles and risk have emerged. The risk topics that AI governance theorists espouse align well with the general framing of ESG risk, but AI governance approaches are — with some exceptions — still pretty light on operational approaches including guardrails, measures and

Table 1: AI-Related Risk Posed by Developers versus Deployers (Example Questions)

AI/Machine-Learning Developers	AI/Machine-Learning Deployers/Users
1. What were the ethical and legal considerations that guided algorithmic development? 2. What level of education and experience do the machine-learning programmers/developers have? 3. What data was selected to train the algorithm and why? 4. What approach to algorithmic training was used? 5. How have humans been involved in evaluating and confirming the machine-learning model? 6. How are customers trained in the use and ongoing deployment of the machine-learning model? 7. To what extent can the firm explain how the machine-learning model makes the decisions that it does? 8. Has the organization ever been found to be non-compliant with legislation and/or regulations relating to the use of AI in any jurisdiction where it operates? 9. Has any customer of the organization been found to be non-compliant and/or subject to a lawsuit in relation to its use of the AI supplied by the organization, where the source of the problem has been identified as the purchased product? 10. How much energy does the organization use annually to power the computers that train the models?	1. Should there be a new AI/machine learning-specific standard within social capital addressing an individual's right to know whether they are subject to a human-led versus machine-driven decision? 2. Are the current disclosure requirements relating to data security and data privacy sufficient in an AI/machine-learning world where broader governance considerations like source, quality, accuracy, consent and the integrity of whole-process management play a role? 3. Do existing disclosure topics such as human rights, data security and data privacy need new accounting metrics that are specific to AI and machine learning (i.e., there are considerations for data management with AI that are incremental to other data applications based on the need to train, finalize and then run models)? 4. Should an organization that employs machine learning report its approach to monitoring model drift and incidence of drift? 5. Should organizations using machine learning report where it is being used and how they have modified their governance approaches as a result? 6. Do members of the public always have a right to know when AI/machine learning is being used and when they are subject to a decision made by a machine?

Source: Authors.

metrics. We have draft guidance from the ISSB on how a global standard will be built for disclosure of sustainability-related risks and opportunities, where there is inadequate consideration of AI-related risks. Based on the general requirements in the draft guidance and the industry-specific guidance in the SASB standards, we have a template from which to work to integrate AI risk-related concerns into ESG reporting and disclosure.

The sixth section of the paper explores how what we know about AI-related risk and AI governance can be laced into the evolving guidance on sustainability-related reporting

and disclosure in support of investor needs and long-term enterprise value creation.

Integrating Material AI-Related Risks into ESG

This paper addresses the challenge of integrating AI-related risks into ESG reporting by taking the ISSB's draft guidance as a template (including the SASB standards) and exploring what salient pieces of AI governance practice might be layered in as inputs, and how this might be accomplished. The purpose of this exercise is the promise of a more structured and standardized global approach to AI governance, both principles and practice. This global baseline should be constructed to facilitate the operationalization of a robust AI governance practice through:

- the establishment of enterprise-level and use case-level policies, processes and procedures, with associated metrics/targets; and
- the establishment of an accountability regime, including clear roles and responsibilities for AI governance processes within the entity.

The discussion below suggests an approach to developing standards for AI-related, sustainability-related disclosures that embraces both general disclosure requirements (i.e., based on the ISSB draft guidance) and specific disclosure requirements (i.e., based on the format of industry and topic-specific disclosure requirements as found in the SASB standards). The decision to work with both general and specific disclosure requirements reflects the guidance from the ISSB, but also reflects the reality that different companies and industries have varying roles, engagement and intensity with AI systems.

A series of preliminary recommendations are made for an approach to integrating material AI-risks into ESG reporting frameworks. These recommendations will be of interest to anyone with an interest in trustworthy AI, AI governance and ESG reporting, but are specifically aimed at the working groups of the ISSB, with a view to building out the evolving global sustainability standards to include AI-related risk and opportunity. While the focus of this paper has been on the integration of AI-related risk into ESG reporting, the authors acknowledge that both AI-related risk and opportunity issues are material to investors and stakeholders. In making these recommendations, the authors therefore parenthetically acknowledge AI-related "opportunity" despite not having delved into the nature of that opportunity in any fulsome way.

- **Recommendation 1:** Sustainability-related risks (and opportunities) relating to organizations' development, procurement and/or use of AI should be included in the efforts of the ISSB to deliver a comprehensive global baseline of sustainability-related disclosure standards that provide investors and other capital market participants with information about companies' sustainability-related risks and opportunities to help them make informed decisions.
- **Recommendation 2:** In approaching the integration of AI-related sustainability-related risks (and opportunities) into the global baseline for sustainability reporting and disclosure, consideration should be given to both general disclosure requirements and industry-specific/topic-specific risks (and opportunities).
- **Recommendation 3:** As part of the process of undertaking integration of AI-related sustainability-risks (and opportunities) into the global baseline for sustainability reporting and disclosure, a definition of AI should be constructed through consultation with stakeholders.

General Disclosure Requirements

A starting point for the development of requirements for general disclosure, is to consider the risks (and opportunities) of AI for each of the four core content areas provided in the ISSB draft guidance: governance, strategy, risk management, and metrics and targets. There is a challenge, however, in a blanket approach to general disclosure on AI, in the absence of organizational context. For this reason, any assessment of the four core content areas should be accompanied by an Organizational AI Statement of Context, clarifying the organization's role and intensity with respect to AI systems. The content of the statement of context might borrow from or even reference the ISO 42001 Context of the Organization standard. As an example, the material topics for disclosure may be expected

to vary with the role(s) of the organization with respect to AI systems, including AI development for own use, AI development for use by others, AI procurement for own use, AI procurement for use by others, platform hosting AI tools for own use and platform hosting AI tools for use by others.

- **Recommendation 4:** AI-related, sustainability-related general disclosure should include an Organizational AI Statement of Context regarding an organization's role with respect to AI systems.

- **Recommendation 5:** Consideration should be given to the construction of guidance such that an Organization's Statement of Context will influence type of disclosure required relating AI-related, sustainability-related risk (and opportunity).

The extent to which core content questions are relevant is likely to hinge on the role and intensity of the organization with respect to AI systems and the anticipated opportunities, risks and impacts of their deployment. In organizations that are light users of AI, for example, it may make sense to report and disclose on AI-related risks and opportunities in the general basket of the organization's sustainability-related disclosures. On the other hand, an organization that is intensely active in AI development and/or deployment, or involved with high-risk, high-impact AI implementations, should be encouraged to undertake AI-specific sustainability-related disclosures due to the novelty and potential impact of the technology.

- **Recommendation 6:** Consideration should be given to the establishment of a tier- or threshold-based approach to general disclosure requirements in relation to AI-related risks and opportunities, reflecting the merits of proportionality and the desirability of avoiding unduly burdensome reporting on organizations who use only low-risk or low-impact AI.

With the Organizational Statement of Context in hand, the types of sustainability-related information that might be considered material for disclosure on governance, strategy, risk management, and metrics and targets can be fleshed out. Borrowing from the ISSB draft guidance, a range of questions querying AI-related, sustainability-related material information should be developed. Examples of the types of questions that may be relevant for each core area are found in the Appendix, and reflect the reality that AI use cases are unique from one to another, and tend to unfold at the sub-enterprise level, within discrete business units.

- **Recommendation 7:** The ISSB Draft Guidance for General Requirements for Disclosure of Sustainability-related Financial Information should be evaluated with a view to adapting and expanding on the questions and topics to reflect the breadth of material AI-related, sustainability-related risks (and opportunities).

Industry-Specific Requirements

In recognition of the fact that the opportunity, risk and application of AI will vary across industries, and that there may be material differences in particular between users and developers of AI, it is recommended that the development of future global standards for AI-related disclosure contemplate the industry-specific approach of the SASB standards.

- **Recommendation 8:** The SASB standards should be evaluated with a view to establishing their adequacy in covering the questions and topics that reflect the breadth of material AI-related, sustainability-related risks (and opportunities) to the extent that these are:

  - not likely to be well-covered through general disclosure requirements;

  - likely or suspected to include different material risks (and opportunities) across different industries; and

  - likely or suspected to include different material risks (and opportunities) depending on whether the organization's role in AI is predominantly as a developer or a user or both.

Considerations for a Path Forward

For organizations that are developing or using AI and machine learning, or doing both, there are good reasons to consider the special governance challenges that come with it. The requirement

for accountability, transparency and fairness relating to AI is a growing public expectation, and is becoming a legal requirement in some jurisdictions. These are material questions for firms and their stakeholders, including investors. The question of how organizations should report and disclose on material information pertaining to their AI-related risks and opportunities is complex, tapping regulatory, legal, compliance, ethical and public relations considerations.

ESG reporting provides a thoughtful approach for how to assess material risks and opportunities relating to AI and machine learning, but does not yet open the door practically for organizations to measure, monitor and report on its use. None of the prevalent ESG frameworks today incorporate standards specifically designed for disclosure relating to AI. This paper aims to stimulate that possibility, with specific intent to engage the ISSB and its working groups in consideration of the requirements for AI-related, sustainability-related reporting and disclosure.

The recommendations set forth in this paper have been developed in recognition of the ISSB's momentum in establishing a global set of standards around reporting on the ESG impacts of firms, and the relevance of these impacts to the investors of firms that develop and deploy AI. The emphasis on ESG disclosure that is financially material to investors aligns with the current direction of the ISSB but does not deny the merits of ESG disclosure more broadly. Financial materiality speaks to the materiality of an individual firm's ESG information in terms of impact on future cash flows and, therefore, the value of the enterprise to an investor. In addition, there is increasing recognition in the role of beta information, or how a firm's ESG practices impact the costs that a firm externalizes to the economy, which, in turn, affects overall securities market returns (Alexander 2022). Without beta-related information, disclosures fail to capture the extent to which one firm's practices impact the returns of other companies in an investor's portfolio and across the economy as a whole. Beyond this, there is a call for ESG-related disclosure that does not affect investors, but is relevant to the impact of firms on other stakeholders.

The additional vectors of information referenced above represent opportunities to build out AI-related sustainability reporting, but the authors argue there is value in starting with the initial guidance of the ISSB and disclosure based on financial materiality. There is an opportunity to advance AI-related measurement, reporting and disclosure by merging select content from the world of AI guidelines and governance into existing ESG reporting frameworks. Today there is no standardized approach to reporting and disclosing on an organization's AI-related strategy, activity, performance, risks and/or impacts. While significant growth in the number of principle-based ethical AI frameworks has been witnessed over the last five–six years, the movement to operationalize them with practical governance approaches is neither robust nor widespread. Recognizing the critical juncture that sustainability reporting has arrived at — on the verge of realizing a single set of global standards — this paper is focused on the merits of integrating topics of AI governance into those standards. The merits of this integration may be found both from the establishment of a global set of standards around AI governance and in the promise of those standards containing operationally focused content.

Works Cited

Alexander, Frederick. 2022. "One Small Step from Financial Materiality to Sesquimateriality: A Critical Conceptual Leap for the ISSB." Harvard Law School Forum on Corporate Governance, May 4. https://corpgov.law.harvard.edu/2022/05/04/one-small-step-from-financial-materiality-to-sesquimateriality-a-critical-conceptual-leap-for-the-issb/.

Brusseau, James. 2023. "AI human impact: Toward a model for ethical investing in AI-intensive companies." *Journal of Sustainable Finance & Investment* 13 (2): 1030–57. https://doi.org/10.1080/20430795.2021.1874212.

Davenport, Tim. 2019. "AI Is Destroying Traditional Business Thinking." *Forbes*, April 5. www.forbes.com/sites/tomdavenport/2019/04/05/ai-is-destroying-traditional-business-thinking/?sh=56fb27993205.

Governance and Accountability Institute. 2021. *Sustainability Reporting in Focus: S&P 500 & Russell 1000 Companies.* www.ga-institute.com/index.php?id=9128.

IBM. 2022. "IBM Global AI Adoption Index 2022." www.ibm.com/watson/resources/ai-adoption.

IFRS. 2022. "IFRS S1 General Requirements for Disclosure of Sustainability-related Financial Information." www.ifrs.org/issued-standards/ifrs-sustainability-standards-navigator/ifrs-s1-general-requirements/#about.

McKinsey & Company. 2022. "The state of AI in 2022 — and a half decade in review." December 6. www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review.

Ocean Tomo. 2021. "Intangible Asset Market Value Study." www.oceantomo.com/intangible-asset-market-value-study/.

Sætra, Henrik Skaug. 2021. "A Framework for Evaluating and Disclosing the ESG Related Impacts of AI with the SDGs." *Sustainability* 13 (15): 8503. https://doi.org/10.3390/su13158503.

———. 2022. "The AI ESG protocol: Evaluating and disclosing the environment, social, and governance implications of artificial intelligence capabilities, assets, and activities." *Sustainable Development* 31 (2): 1027–37. https://doi.org/10.1002/sd.2438.

SASB. 2018. "Software & IT Services." Sustainability Accounting Standard.

Silverman, Karen. 2020. "Why Your Board Needs a Plan for AI Oversight." *MIT Sloan Management Review*, October 19. https://sloanreview.mit.edu/article/why-your-board-needs-a-plan-for-ai-oversight/.

Appendix

Example Questions for Disclosure Topics of AI Risk-Related Information: Based on Draft Guidance Contained in the ISSB's Exposure Draft (S1) General Requirements for Disclosure of Sustainability-Related Financial Information

Governance-related disclosure (speaks to accountability and process within the entity for sustainability-related risks and opportunities):

- Is one overarching body responsible for oversight of AI-related risks and opportunities within the entity? If so, please identify the body.
- How are the responsibilities of the body(ies) captured in terms of reference and policy?
- How do(es) the accountable body(ies) ensure appropriate skills and competencies?
- How do(es) the body(ies) consider AI-related risk and opportunity in the context of strategy and risk management including trade-offs?
- How do(es) the body(ies) establish targets for AI performance including on ethical grounds and monitoring of progress?
- What is(are) the body's(ies') description of the role of management in all this?
- How do(es) the body(ies) responsible for oversight of AI-related risks at the enterprise level, delegate or trickle down responsibilities and accountability for AI programs and projects at the use-case level?
- How do(es) the body(ies) charged with oversight of data governance and AI-related risks and opportunities function together?

Strategy-related disclosure (speaks to what sustainability-related information may be material in relation to an impact on business model and financials):

- What AI-related risks and opportunities face the entity that it reasonably expects could affect its business model, strategy and cash flows, its access to finance and its cost of capital, over the short, medium and long term? And how does the entity define short, medium and long term?
- Given an entity's position on the AI-related risks and opportunities it faces, what might be the effect of these on its business model and specifically its value chain?
- Given an entity's position on the AI-related risks and opportunities it faces, what might be the effect of these on strategy and decision making?
- Given an entity's position on the AI-related risks and opportunities it faces, what might be the effect of these on financial position, financial performance and cash flows?
- Given the understood AI-related risks and opportunities, what is the resilience of the entity, or its capacity to adjust to the uncertainties arising from these risks?

Risk management disclosure (speaks to details of how the sustainability-related risks and opportunities are identified, assessed and managed in order to enable evaluation of the entity's risk profile and risk management processes):

- What is the process by which the entity identifies AI-related risks and opportunities for risk management purposes? Is the entity's approach to AI-related risk management implemented at the enterprise level or the use-case level or both?
- What is the process by which the entity assesses, prioritizes and monitors AI-related risks and opportunities? Is the entity's approach to the assessment, prioritization and monitoring of AI-related risk implemented at the enterprise level or the use-case level or both?
- To what extent is the AI-related risk management process integrated into the enterprise's overall risk management process?

Metrics and targets disclosure (speaks to how the entity measures, monitors and manages its significant sustainability-related risks and opportunities in order to understand how the entity assesses its performance, including progress toward targets it has set):

- The entity is to include metrics (enterprise level and use-case level) that apply to the activities in line with its business model and in relation to AI-related risks and opportunities. There may be different sustainability-related risks and opportunities that are driven by AI across different businesses/industries in which the entity is engaged.
- The entity is to disclose the metrics (enterprise level and use-case level) it uses to manage and monitor AI-related risks and opportunities and performance, including progress against established goals and targets.
- The entity is to disclose details relating to the development of the metrics (enterprise level and use-case level) relating to AI-related risks and opportunities, including how it was defined, whether a third party validated it and what methods were used to calculate the targets.
- Special consideration should be given to the establishment of metrics that pertain to enterprise-level risks and opportunities (for example, transparency of AI governance and processes, quality of talent, access to talent, quality management system performance, risk management system performance, compliance record, media comment, stakeholder feedback) versus metrics that pertain to the use-case level, which, in turn, could be aggregated (for example, AI system performance, AI system fairness, AI system explainability, AI system security, data security and privacy in the context of the AI system).

Centre for International
Governance Innovation

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org

@cigionline