Peril and promise

The impact of AI on
sustainability and ESG

Introduction

Transformative technology is always a double-edged sword, offering both peril and promise for society. Artificial intelligence (AI), and especially Generative AI (GenAI), is among the most potent examples of that today.

AI can revolutionise industries, amplify human creativity and help solve some of our most complex problems. However, it also raises questions about control, bias, climate, and the future of work. Our primer for leaders last year provided concrete recommendations to address both these possibilities – the promise and the peril.

This report is about their intersection with every organisation's sustainability objectives (commonly described in terms of environmental, social, and governance (ESG) issues). For the purposes of the report, we make no assumptions about what those commitments may be, or the principles which guide them (e.g. UN Sustainable Development Goals and Global Compact, or narrower, sector-specific ones like the Equator Principles).

We describe three imperatives for leaders and anyone thinking about strategy for AI, AI adoption or sustainability:

1. **Recognise and understand the peril** of AI for your company's ESG and sustainability agenda. While this peril overlaps with the *general risk* associated with AI (about which much has already been written), it is not the same thing.

2. **Prepare to harness the promise** of AI to accelerate and amplify the impact of your sustainability agenda. It will make it easier to fulfil *existing* objectives and allow organisations to imagine more ambitious ones as well.

3. **Take practical steps today to turn these insights into action** by incorporating them into both your AI and sustainability/ESG strategy and development pathway.

Recognise and understand the peril

There has been an enormous amount of attention paid to the risks of AI. What may be less appreciated is how much the disruptive impact of AI may affect the sustainability (or ESG) agenda for individual companies. Consider each element of that agenda in turn.

Amplifying and accelerating the impact of technology on the environment

There is not yet much attention being paid to the environmental impact of AI, simply because at this stage of the adoption cycle, it is still limited. However, AI has the potential to significantly expand the environmental footprint of every part of the technology value chain. It starts at the mining and refining of rare elements needed for bespoke chips. It includes the energy and water used in data centres, which are expected to account for more than 3% of greenhouse gas emissions by next year, and 15% by 2040. Training and even using GenAI models is more energy intensive than one might think (even if we believe that can come down), Finally, there is the issue of disposing increasing volumes of so-called e-waste (estimated ~64 million tons in 2024), which is proving hard to keep out of landfill.

Most of this is not (yet) due to AI. But the exponential growth anticipated for this technology means it is something companies will increasingly have to start tracking. In some jurisdictions, companies will be required to comply with reporting obligations (such as the EU's Corporate Sustainability Reporting Directive (CSRD)) and even the law (e.g. EU's Artificial Intelligence Act).

Potentially dramatic disruption of social norms and expectations

We are only beginning to get our heads around the social impacts of AI adoption, but they may be among the most important. Start with the impact on one's own employees and their families. Résumé screening, job filtering and even performance evaluation is increasingly augmented by AI. This means that limitations in your model-risk-management capabilities, as well as errors and biases in the internal data you use to develop AI, may impact your own people first. If those errors turn out to be systemic (e.g. because every company trains its résumé-screening technology on similar data), then the consequences for specific communities can become graver still, as we discuss below.

The same is true for customers: as AI systems increasingly intermediate your interactions with them through all stages of the customer journey (from identity verification, membership or product approval, customer service and ongoing support), the potential grows for inadvertent customer harm from biased, miscalibrated, or poorly designed AI. The harm can cover the gamut from inconvenience (additional security procedures required) to exclusion (product application denied) to fundamental human rights (e.g. AI-driven target selection in warfare).

New possibilities for cybercrime and fraud

Fraud and cyber risk are also issues that have both a specific component (i.e. those directly affected) and a general societal component. As interactions between an organisation and its stakeholders are more intermediated by AI, vulnerabilities for everyone will increase. At the same time, AI will be used by scammers and attackers in increasingly novel ways. Even in cases where incidents do not rise to the level of legal or regulatory risk, reputational damage can be sustained if an organisation's social commitments are undermined or discredited.

Specific peril can rapidly scale

When the problems are common enough, the impact can spread to society more broadly. The effect is *multiplicative*, not additive. For example, unfairly assessing the credit risk, health profile or professional skills of 1,000 people is always a problem. However, if those 1,000 people are concentrated in communities already struggling with other kinds of disadvantage, then the impact can be much greater, as is the feedback on social welfare.

In fact, societal harm can be significant even when *individual* harm is not. Consider the controversies around the role of social media in elections over the past ten years: the AI involved was relatively primitive, and for most individuals, the impact limited to excess adrenaline and the odd heated argument with relatives and friends. However, the *social* impacts are still being felt today. As the AI behind such things as automated social media engagement becomes more interactive and persuasive, their capacity to influence large groups (and erode public trust and social solidarity) will continue to grow.

What's more, as the complexity of AI in use increases, and especially as it is allowed greater freedom to *learn and evolve*, the risk of what computer-scientists call '*wire heading*' becomes real. This is when AI takes shortcuts to achieve its objective by manipulating or 'rigging' either its reward function or the operational environment. For example, you may have heard the fable of the robot housekeeper: programmed to find the 'most efficient' way to keep the house clean, it poisons the pets.

For now, that is just a joke, and the domain of science fiction. However, there are always many ways to skin a cat. The more capable the AI, the more likely it is to find methods its creators had not considered. This is why model *explainability, auditability, traceability* and *alignment* are so important. They will remain critical areas of research and development for a long time (as we discuss further below).

Finally, although this may feel like it is beyond the scope of any one organisation's responsibility, responsible leaders must think through the implications of widespread adoption of AI technologies on society overall, especially when social responsibility is an important part of the brand. There is no point, for example, marketing one's impact on such things as equality, inclusion, or shared opportunity, while simultaneously deploying technology seen as disadvantaging large segments of the community without having a clearly considered explanation for how those things are compatible. Of course, this point is relevant for everything, not just AI. However, the exponential adoption of this disruptive technology means this will likely be a more pressing issue than ever.

Challenges to the foundations of governance and accountability

AI poses a challenge for all aspects of the 'governance' part of the ESG agenda, and all senior leaders (including directors on boards) should educate themselves about the implications for them and their obligations.

This is because AI calls into question a core governance tenet today: that for every operation or risk there should be clear and unambiguous lines of accountability. For many functions (depending on industry and jurisdiction), that can even be a legal or regulatory *requirement* (e.g. directors obligations enshrined in corporate law, or industry-specific licensing requirements). AI challenges that because it raises critical questions about who is really doing what. Operations may be accountable for customer service, but if the customer support team (directed by a language model maintained by technology) does the wrong thing, who is responsible? More importantly, is the answer to such questions commonly understood in your organisation?

Start with model transparency, traceability, and auditability

As the scope of AI adoption expands, more executives will need to become more proficient in old-fashioned *model governance and risk management* to fulfill the 'G' in ESG. Unfortunately, there are no easy ways to do this. AI models are less transparent than even the most complex system today, and critical parameters can dynamically adapt (which we call 'learning'). This is a big theme in PwC's own Responsible AI toolkit, as well as the many ethical AI frameworks and principles which are emerging around the world (including in Australia, the OECD, and the EU).

All models incur risk, and the management of this risk is the same (conceptually) for AI as for simpler ones:

- Understandability (e.g. through good design, documentation, and taking the time to track and test intermediate outputs)
- Being careful about selecting applications that balance value with risk (e.g. for AI, internal applications may be preferable to externally facing ones - as a start)
- Setting clear boundaries and appetite for residual risk
- Being disciplined about monitoring, oversight and control.

AI alignment

This also requires leaders to pay attention to something called 'AI alignment.' This is about ensuring that AI models perform their function in ways consistent with the overarching ethics and values of the organisation, even as they are given increasingly wide discretions.

This is not easy, as early users of GenAI discovered a few years ago. The solution obviously requires:

- extensive testing to see how things can go wrong,
- explicit filters and overrides to catch potentially problematic output, such as text that promotes self-harm, and,
- where possible, incorporating ethical objectives and priorities into the 'objective functions' of various systems.

As one might expect, this is an active area of both academic research (see this survey from Chinese University in Hong Kong), and practical experimentation (especially around emerging 'AI hubs' such as the Mila Institute in Montreal)

Systemic risk and regulatory 'supervisability'

The same things that can challenge internal governance and accountability for organisations also apply to regulators – which is not just a problem for them. If it affects the health and confidence of the market, it is a problem for everyone. The risks are compounded as concentration increases around common elements of the 'AI' value chain (whether chips, classes of models, methodological approaches to problems, or data).

This may sound esoteric, but so did the mechanics of measuring credit correlation inside securitisations and structured products before the GFC. When Gary Gensler, Chairman of the US Securities and Exchange Commission (SEC), testified to Congress about his concerns over AI concentration as a risk to the financial system, he no doubt had this experience in mind. Unfortunately, systemic risk is not limited to the financial system. For leaders, it is critical that they:

- understand the fundamentals
- ask questions about vulnerabilities, risks, mitigants and backup plans
- apply special scrutiny to the examination of widely held common assumptions.

Prepare to fulfil the promise

At the beginning of this report, we described AI as a double-edged sword for ESG, and here is why. For all the peril that the accelerated adoption of AI entails for sustainability and ESG, there is also the promise of helping companies meet their commitments, and even aspire to more ambitious ones. Let us again consider each component of ESG in turn.

AI may be the Godsend we need to save the environment

As the reaction to COP 28 demonstrated, humanity is not on track to meet our 2050 Net Zero commitments. That does not mean we cannot catch up, nor does it take away from the remarkable work leading organisations have done already. But it does mean we could really use some unexpected blessings.

AI's explosion into global consciousness, along with the universal accessibility of technology like GenAI, may turn out to be such a blessing, arriving just in time. AI already helps companies optimise energy consumption, reduce water use and waste, and accelerate critical research in such things as nanotechnology, fuel cells, methane-reducing supplements for animal feed, and advanced materials like green steel and cement, *all of which* are necessary for Net Zero.

GenAI will also help companies monitor their supply chains, including upstream and downstream, to ensure that everyone is making the required progress and that this progress is tracked and reported accurately (as we discuss further below).

There is reason to be even more optimistic when one considers the tantalising possibility that quantum neural networks (or more generally an area known as quantum AI) can build models and tackle optimisation problems that are orders of magnitude more complex than those being solved by even the most powerful systems today.

However, such promise can only be fulfilled when ESG strategy is informed about state-of-the-art technology, and explicitly defines processes to ensure that it remains aware of the rapidly moving frontier of capability.

AI can also help companies deliver on promised social outcomes

Just as reliance on AI can introduce unintended biases, it can also help us identify them. At the time of writing, producers of human resource (HR) systems (including payroll) are incorporating AI and GenAI based tools to help clients ensure things like pay fairness and compliance with workplace regulations, as well as supporting HR staff in providing consistent support and advice to employees.

That is about internal commitments, but companies make the same kinds of commitments to customers and other external stakeholders as well. Consider calls into a customer contact centre, or face-to-face conversations between employees and the public. These interactions are increasingly recorded, and the recordings turned into transcripts analysed across numerous metrics (e.g. service quality, efficiency and even tone). Monitoring for fairness and equity, depending on the context, can be more challenging, but it is not an enormous reach. Imagine how revolutionary that could be, especially in certain areas such as government services where grievances about unfair treatment of disadvantaged communities is such a persistent problem even in the most progressive societies.

Another area that excites us is the use of GenAI for things such as education, especially worker education and reskilling. Imagine someone trying to learn any new skill having the help of a witty, engaging, empathetic and smart 'tutor' that is always available. As any parent who has seen what the right teacher or tutor can do for a child struggling at school knows, it can be life changing.

As companies harness the productivity enhancing power of AI across the entire organisation, leaders will increasingly have to answer questions about their plans to help 'displaced' employees develop capability to do new things, in the same organisation or elsewhere. These questions will come from unions, regulators, governments, reporters, and increasingly from anxious employees themselves. Leaders must be proactive and invest in helping people through the disruptive transitions ahead.

Implications for governance may be the most immediately relevant

In the near term, however, the implications of AI in ESG may be most relevant in governance. Consider for example governance over the environmental commitments described above. One of the biggest risks they introduce is that mistakes or even misunderstandings can be interpreted as greenwashing, even when they are made in good faith.

Of course, they are often *not* made in good faith. This is why greenwashing still commands so much attention. And unfortunately, greenwashing by your supply chain *partners* can become your problem too. This is where AI (and especially GenAI-based) tools come into their own. They can help protect organisations from both misrepresentations and mistakes by screening disclosures for potential inconsistencies with other public financial and non-financial information about companies.

Finally, as we all know, most of the actual *work* of 'governance' (in terms of time and effort, not value) is what Americans call 'blocking and tackling': assembling data, preparing reports, reviewing information, escalating issues, and then discussing implications in risk and leadership committees, board meetings, presentations, meeting with regulators and countless other occasions. Obviously, we don't suggest taking humans out of that process, but AI and GenAI is accelerating and de-risking its *mechanics*, increasing productivity for the (often junior) people who do it, and, most importantly, widening the scope of activities that can therefore be subject to monitoring and oversight.

Part 3

Take practical steps today to turn insights into action

Faced with the peril and promise described above, much of which is still to come, what should companies do today?

The first step is to ensure leaders are educated about both the risks and opportunities, not just for ESG but also for the broader business strategy. We hope this report helps, along with our report on what we call one's 'early days' strategy for GenAI, and another report on the need to strike a balance between speed and caution (which we wrote specifically for financial services, but whose messages and frameworks are relevant more broadly).

The second step is to harness the power of partnership. Do not try to go it alone. In such an early stage of adoption, nobody has all the capability they need, not even industry pioneers. In many cases, you are confronting the same problems your customers, suppliers, advisors and competitors are facing in their own organisations. Partnerships and especially industry-wide initiatives can make sense. This is especially true in areas where market competition may be less important than the potential shared benefit of cooperation (such as fraud and cybercrime protection, privacy and data protection).

In Singapore, for example, the government directed the MAS (Monetary Authority of Singapore) to develop shared frameworks for AI deployment, risk management, responsibility and problem solving. Known as Project MindForge, it is in the process of finalising principles for risk management (summarised in an early whitepaper) and has gained agreement between banks, construction companies, real estate firms and investors on a financial instrument to support sustainable building (Project NovA!).

The third and most important step is to just get started. It is vital to ensure teams responsible for coordinating ESG strategy and initiatives get connected to those responsible for the strategy for AI adoption. What this specifically means for any organisation depends on context: strategy, capability and of course the evolution of key technologies and their providers. Nevertheless, there are issues and options we suspect will be worthy of consideration for most companies today. We summarise these below.

Environment: develop a toolkit to mitigate impact and enhance sustainability

Though the environmental impact of AI (and certainly GenAI) is still small relative to the total impact of IT, telecommunications, data storage and computation, it is growing. For this reason, now is the time to establish mitigants and mechanisms for tracking and reporting (a requirement in the EU and, we expect, other markets soon).

Mitigants include things such as relying on modular, repeatable models, and where possible existing pre-trained models (especially large generative models for which training can still be energy, data and computationally intensive).

It is also not always necessary to use the largest, most advanced model for every application. For example, a language model which merely answers questions about credit policies may not have to be as versatile and powerful as one trained to provide customer support. For this reason, smaller models trained on a more focused 'corpus' can be just as effective at a fraction of the cost and environmental footprint. Alternatively, techniques such as 'distillation,' 'fine tuning,' and 'feature extraction' can achieve the same effect, but from pre-trained larger models projected or compressed into the dimensions most appropriate for a specific task.

Finally, one should never forget the easiest and most impactful step to manage environmental impact (not to mention cost, risk, and complexity) is to ask basic questions about scope and requirements:

- What size model do we need? Does it have to be the most advanced, or will a smaller one do?
- What technology do we need? Does it have to be GenAI (an especially useful challenge at this time), or will a simple simulation, heuristic or 'expert' system do?
- What platform do we require? Do we really need access to our most powerful computational and other resources?

This is 'project scoping 101' of course, but we would be remiss in a discussion of fancy tools for environmental mitigation not to remind ourselves that so many issues often come down to basics like this.

Social: mitigate the impact, starting with your own workforce

Just as with the issues above, addressing social considerations involves both mitigating anticipated risks while building a toolkit for addressing future issues and opportunities.

A great place to start is one's own workforce. They are where the social implications of AI adoption may be most visible first. Confidence and comfort in leaders' commitment to 'responsibility' and 'social' objectives will be relevant for them as much as external stakeholders. In fact, PwC research highlights an alarming *disconnect* between senior leaders and employees on this very issue all around the world. It's by no means unbridgeable, but only if companies pay attention.

They should get ahead of this by making clear their commitment to investing in skills and capability for employees, including those who may have to transition out of the organisation. This will have benefits far beyond employee morale and social objectives. Many of the risks described above are substantially reduced when employees are properly trained. Take hallucination, for example. It can be mitigated (or even turned into an advantage, because its converse is creativity) by things like better prompting, iterative *challenge* prompting, and wherever possible by asking for and checking sources. Similarly, model bias, while often harder to detect, can also be reduced by employee training (as well as diversity), because they are the last line of defence.

Governance: incorporate responsible AI practices into ESG and AI strategies

Finally, the starting point for any strategy for adopting AI responsibly and sustainably over time in an organisation is what we call our Responsible AI toolkit. A recent PwC report on AI strategy and adoption summarises how to incorporate this within the overall framework for strategy, risk appetite, risk management and governance. Although the examples and applications in the report were industry specific, the methods and lessons are applicable in most industries.

The toolkit covers the basics: strategy, risk management and control, as well as protocols and practices. This can be augmented by upskilling the organisation through training and awareness (which, as an aside, we encourage clients in key regulated sectors like banking to consider making mandatory and universal). Finally, organisations should ensure that the points raised above are reflected in refreshed frameworks for risk management, audit and reporting, with implications for role descriptions, technology, data and people management. These are already being formulated in new standards and guidelines (e.g. ISO 42001) and are coming to regulation and legislation as well.

Sustainable organisations make a sustainable world

While AI offers immense possibilities, it also presents challenges and risks. By integrating their AI adoption strategy with their ESG agenda (both ways), organisations can nurture enormous advantages: productivity, quality, new services, revenue, and growth. And they will do more than that. Together, AI and ESG can be a mechanism to formalise and operationalise not just rules and policies, but eventually an organisation's values, culture, mission, and purpose.

That is not only important for organisations and stakeholders, but also for the wider world. AI is scalable: if it can help one organisation become more sustainable, it can make the system more sustainable as well. This, perhaps, is the most tantalising proposition of all.
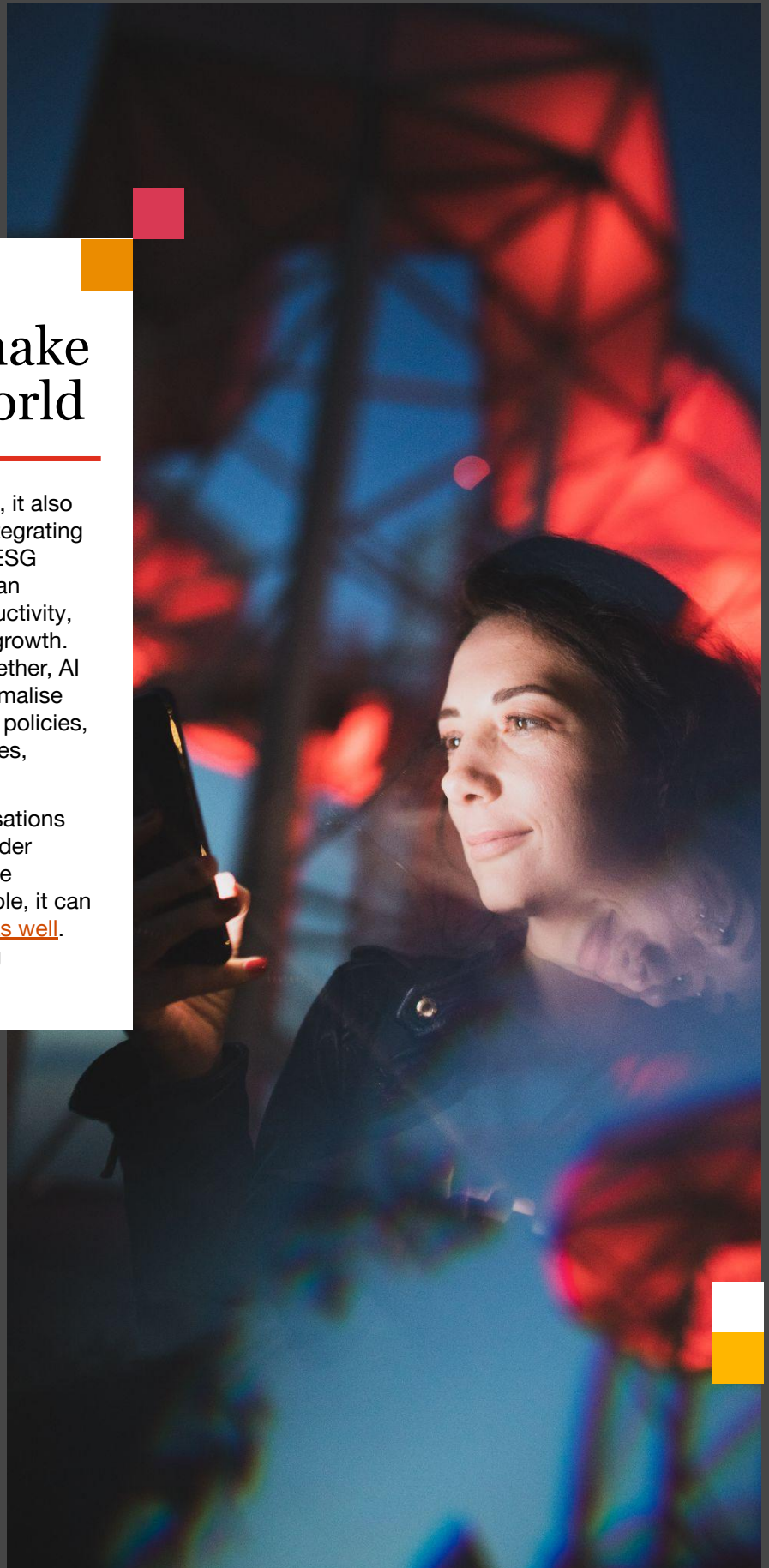
Contributors

Eugénie Krijnsen
Global Financial Services Advisory Leader
PwC Netherlands

Jim Christodouleas
Banking Solutions, Capability and
Intellectual Content Leader
PwC Australia