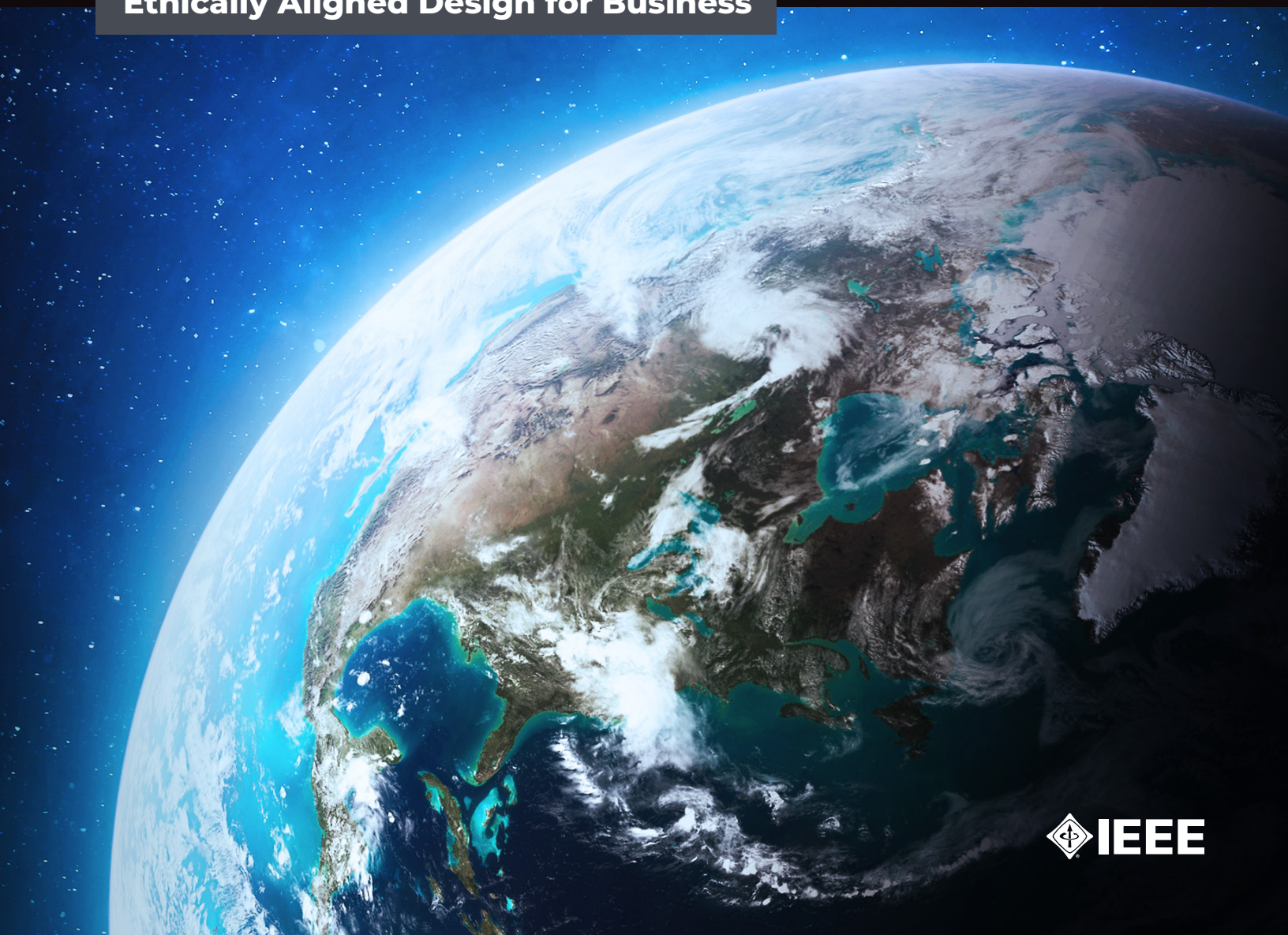
IEEE SA STANDARDS ASSOCIATION

Prioritizing People and Planet as the Metrics for Responsible AI

Ethically Aligned Design for Business



IEEE

Introduction

You may be aware of the term responsible artificial intelligence (AI). Perhaps you are already using a methodology for your AI design based on European regulations to avoid risk. Or you are working to help ensure algorithms are not biased to avoid unintended consequences. No matter your specific situation, defining how to be responsible with artificial intelligence systems (AIS) is critical for modern technological design.[1]

In this paper, we will explore the question: What are the metrics of success for Responsible AI? Our primary goal is to provide direction for business readers so they can utilize these metrics—large enterprises as well as small- and medium-sized businesses (SMBs)—while also informing policy makers of the issues these metrics will create for citizens as well as buyers. We recommend that organizations define what responsible means at the outset of design and throughout the entire life cycle of their AIS by being accountable to expanded key performance indicators (KPIs). While common business performance metrics focus on financial indicators primarily, organizations risk causing unintended harm when issues of human well-being or ecological sustainability are not prioritized in their planning.[2]

It can be tempting to see words like well-being or sustainability and not really consider them metrics of success. Where only financial or growth-oriented indicators are used to demonstrate value, human and planetary flourishing are often not measured. In other words, all issues stemming from human or planetary harm may not be overtly included in the design and implementation of technology. This is the most significant risk we face with artificial intelligence today, not bias or a lack of transparency but thinking that we can ignore the needs of our planet or human mental health for the sake of growth. Just because gross domestic product (GDP) and other financial metrics are universally known does not mean well-being indicators like the Sustainable Development Goals (SDGs) of the United Nations (UN) or other environmental, social, and governance (ESG) standards do not exist—or that they do not have pragmatic ways to measure their use. This is a key reason we are writing this paper: to let you know about these other metrics because they do matter. And they matter more than economic indicators and financial metrics for a very specific reason.

[1] While we use the term artificial intelligence here to honor the historical introduction of the term, we prefer and recommend the term artificial intelligence systems (AIS), as defined by OECD: "An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy."

https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. The term AIS will be used from this point on.

[2] A 2019 report from UNESCO as reported by The Guardian article "Digital assistants like Siri and Alexa entrench gender biases, says UN" shows the need for metrics regarding human well-being. A 2019 study featured in the MIT News article, "Shrinking deep learning's carbon footprint" says "training a large deep-learning model produces 626,000 pounds of planet-warming carbon dioxide, equal to the lifetime emissions of five cars" shows the need for holistic sustainability metrics.

For our definition of well-being, we are using a definition created by expert members of the IEEE Standards Working Group that created IEEE Std 7010™-2020, IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being. Well-being is defined as:

The continuous and sustainable physical, mental, and social flourishing of individuals, communities, and populations where their economic needs are cared for within a thriving ecological environment.[3]

We are going to say something tough to hear: If the science is correct, people may not be around on this planet in 100, 200, or 300 years. The exact timing is both semantic and not the point. The point is that these unidentified people include your children, grandchildren, and the kids of everyone else in the world. If we design technology from this point forward without prioritizing the planet and people, we choose to ignore our future generations out of ignorance or fear. Saying, "it will be hard to change," is a valid excuse but irrelevant. As responsible parents, citizens, or humans, we have no choice but to embrace the metrics that will help ensure the longevity of people, the planet, and purpose-driven profits. So, this is actually really, really good news.

The logic of this thinking is mirrored in the 2020 report from the World Economic Forum (WEF) "Measuring Stakeholder Capitalism: Towards Common Metrics and Consistent Reporting of Sustainable Value Creation."

At the heart of this exercise is the belief that the interrelation of economic, environmental, and social factors is increasingly material to long-term enterprise value creation. But beyond this, those corporations that align their goals to the long-term goals of society, as articulated in the SDGs, are the most likely to create long-term sustainable value, while driving positive outcomes for business, the economy, society, and the planet.[4]

The interrelation of economic, environmental, and social issues must be factored into the definition of responsible AI. This interrelation demands the articulated use of metrics for ESG and UN SDGs (including triple bottom-line considerations of "people" and "planet" as well as "profits") at the outset and throughout the life cycle of AIS.

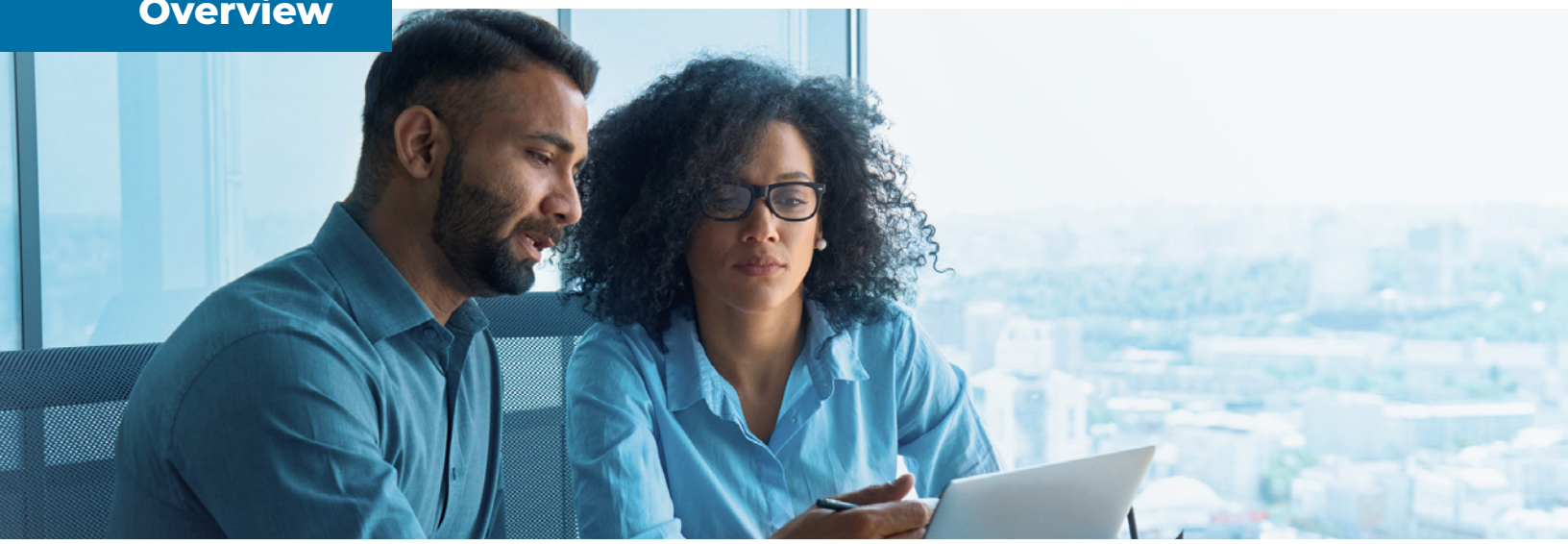
This report provides an introduction and recommendations for how organizations can begin and grow the process of incorporating well-being metrics into creating and maintaining their AIS.

More than a dozen industry experts have provided insights in the following pages based on their front line experiences as responsible AI technologists and AI ethics, ESG/metrics, and sustainability thought leaders and practitioners in their organizations. By relating the recommendations within *Ethically Aligned Design* to their industry efforts, our committee members provide pragmatic applications of responsible and sustainable AI inspired by IEEE's global community of academics, data scientists, engineers, and tech entrepreneurs for inspiration and mutual learning.

[3] IEEE Std 7010™-2020, *IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being,* p. 19. *https://www.weforum.org/reports/measuring-stakeholder-capitalism-towards-common-metrics-and-consistent-reporting-of-sustainable-value-creation.*

[4] World Economic Forum, "*Measuring Stakeholder Capitalism: Towards Common Metrics and Consistent Reporting of Sustainable Value Creation*," Sep. 22, 2022, p. 6, https://www.weforum.org/reports/measuring-stakeholder-capitalism-towards-common-metrics-and-consistent-reporting-of-sustainable-value-creation.

Overview



This paper, "Prioritizing People and Planet as the Metrics for Responsible AI," is an extension of the paper "A Call to Action for Businesses Using AI." The latter was the first paper written by the Ethically Aligned Design for Business Committee, which is part of the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. A key focus of our first paper was that Responsible AI demands human-centric, value-based design. From the introduction:

Ethical decision making is not just another form of technical problem solving. A tech-centric focus that solely revolves around improving the capabilities of an autonomous or intelligent system does not sufficiently consider human needs and the long-term impact on the future of society.

Simply put, if you are working in the world of AI, you are, in fact, working for the future of humanity, so you need to embed ethics practices across all teams responsible for these types of systems. An ethical, human-centric AI must be designed and developed in a manner aligned with the values and ethical principles of the society or community it affects.

Aligning with society's values and ethical principles demands the use and conformance to the metrics for ESG and the UN's SDGs or other well-being metrics described in this paper.

Our first paper featured the following three sections plus an AI Ethics Readiness Framework table:

- The Value and Necessity of AI Ethics
- Creating a Sustainable Culture of AI Ethics
- AI Ethics Skills and Hiring

This paper follows a similar format to feature the following sections plus a Well-Being Metrics Readiness Framework table:

- **Assess:** What organizational values are leveraged for Responsible AI metrics?
- **Apply:** How does an organization implement the principles and metrics of Responsible AI?
- **Amplify:** How does an organization scale its work to grow and measure the success of Responsible AI metrics?

ASSESS:

WHAT ORGANIZATIONAL VALUES ARE LEVERAGED FOR RESPONSIBLE AI METRICS?

"Defining and embedding AI ethics skills is about behavioral change within a wide group of people in a company. It will take time for businesses to consider AI ethics as a core competency across roles instead of only considering it for issues of compliance or in response to negative PR."

—"Creating a sustainable culture of AI ethics," from IEEE's *A Call to Action for Businesses Using AI*

Our first paper in this series provided several key insights into how an organization can embrace concepts surrounding AI ethics. Multiple updates regarding global regulation for artificial intelligence systems (AIS) have taken place since the release of our report, including the Artificial intelligence Capabilities and Transparency (AICT) Act of 2021 in the United States that defines AI ethics as "the quantitative analysis of artificial intelligence systems to address matters relating to the effects of such systems on individuals and society, such as matters of fairness or the potential for discrimination." The fact that the United States recognizes AI ethics as an area to address is of historic consequence.

Likewise, in November of 2021, the United Nations Educational, Scientific and Cultural Organization (UNESCO) released its Recommendation on the Ethics of Artificial Intelligence, which stated, "We need international and national policies and regulatory frameworks to ensure that these emerging technologies benefit humanity as a whole. We need a human-centered AI. AI must be for the greater interest of the people, not the other way around." The Global Partnership on Artificial Intelligence (GPAI) has 25 member countries who, "support and guide the responsible development, use and adoption of AI that is human-centric and grounded in human rights, inclusion, diversity and innovation, while encouraging sustainable economic growth."

"AIS technologies affect human agency, identity, emotion, and ecological systems in new and profound ways. Traditional metrics of success are not equipped to ensure A/IS creators can avoid unintended consequences or benefit from unexpected innovation in the algorithmic age. A/IS creators need expanded ways to evaluate the impact of their products, services, or systems on human well-being. These evaluations must also be done with an understanding that human well-being is deeply linked to the well-being of society, economies, and ecosystems."

—Ethically Aligned Design, First Edition

The European Commission is proposing the first-ever legal framework on AI. The EU AI ACT adopts a risk-based approach, whereby risk assessment goes beyond safety and security and includes risks of violation of fundamental rights, such as dignity. Applications that are in contradiction to these values, like social scoring, will be banned from the European Union (EU) market. Applications in sensitive areas—such as human resources (HR), biometric identification, border control, and access to essential services like healthcare and education—are considered to be high risk, and mandatory requirements will be applicable to the design and development of these AI systems before they are placed on the market.

These trends, along with the growing call for greater due diligence toward environmental sustainability reporting for companies, demonstrate the need for an organization to position its AI (and all of its technologies, products, and services) with a perspective on well-being, taking care of the future through collective stewardship of science and innovation in the present.[5]

While there are multiple ways to implement metrics for Responsible AI, one way to start is with the satisfaction and well-being of your employees. Beyond the need to retain talent in the wake of the Great Resignation, companies are embracing the findings of studies like the one featured in the 2019 Harvard Business Review article "The Key to Happy Customers? Happy Employees," which notes, "There is a strong statistical link between employee well-being... and customer satisfaction among a large sample of some of the largest companies today. A happier workforce is clearly associated with companies' ability to deliver better customer satisfaction.

This research points to a clear logic about how to begin implementing and expanding on Responsible AI metrics in your organization—start with the satisfaction and well-being of your employees.

[5] Stilgoe, Jack, Richard Owen, and Phil Macnaghten, "Developing a framework for responsible innovation," Research Policy, vol. 42, no. 9, Nov. 2013.

Your wellness programs, while likely focused primarily on physical health, provide opportunities to recognize the metrics of success for larger ideals of well-being, which we will explore later in this paper. But where it is understood that happy employees mean happy customers, your Responsible AI lead can start with getting buy-in from your colleagues in areas already established in HR or other departments.

GETTING BUY-IN FROM COLLEAGUES

Companies of all kinds are now in a position to assess what wellness or well-being means to them and how they can incentivize triple-bottom-line behavior going forward—prioritizing people and planet before profit (so finite resources can be cared for in a realistic, long-term sustainability mindset). Notably, some of the categories included in well-being initiatives are not new; supporting sustainable behavior has been a priority for many large organizations over the last several decades. However, some have criticized these efforts as "virtue signaling," more for show than for foundational and strategic organizational decision-making. On the other end of the spectrum are those who believe well-being initiatives put organizations at a competitive disadvantage.

For well-being initiatives to carry weight, they must be tied back to organizational values and principles. Leadership should buy into these principles and should reinforce them. Not all companies are the same; organizations are in different industries, are more or less established, at different maturity levels with well-being initiatives, and have varied funding sources. A company's unique lens is shaped by its customers, investors, partners, and the company itself. These attributes, and more, necessitate an organization-specific approach to defining and refining well-being criteria.

It is critical to recognize that many organizations define well-being as a proxy for employee health or only via perspectives typically handled by areas like corporate social responsibility (CSR). Yet, in the same way that quarterly profits apply to every division and aspect of a company, the metrics encapsulated within the UN SDGs or other established global indicators should be used at the outset of any AI effort for Well-being by Design methodology. This means that anything you build can and should increase the empirical, long-term physical and mental well-being of an end user or customer (while enhancing the health of your employees). This "design with the end in mind at the outset of design" methodology should already be a part of your Responsible AI design practices to avoid unintended consequences of "ethics as an afterthought." This logic applies because human well-being, environmental sustainability, and Responsible AI (once released to the public) relate to the impacts of a product, service, or system that cannot be designed in isolation from a small group of technologists not taking these issues into account.

Start with what you have. Leadership principles, diversity, equity, and inclusion (DEI) initiatives, sustainability initiatives, newly developed ethical data, AI principles, and more provide excellent tools to identify and utilize established well-being metrics like the UN SDGs. In some cases, the values surrounding these tools may exist as a shared ethos but are not written down. Even if a company does not have any principles or values to work from, key company members likely do. Several workshop-based techniques exist (including systems-thinking workshops) to gain stakeholder alignment and uncover a shared view of systems, priorities, and organizational objectives regarding Responsible AI and well-being metrics. Sessions such as these would be useful first steps to defining and refining a set of values that can serve as the basis of a well-being initiative. Outlining and documenting a set of values starts the commitment to employee wellness and enables all stakeholders to align on how well-being metrics can evolve at the company.

Beyond principles and values, existing practices within an organization can be leveraged to pursue wellness or well-being. Compliance functions may use impact assessments (environmental, and human rights, among others) to assess both positive and negative impacts that specific technology, processes, and initiatives may have in the enterprise.

Existing philanthropic efforts, both financial and volunteer-based, can also be a starting point supported by both organizations and their members or employees. Some organizations already supply ESG reporting and diversity and inclusion metrics in their annual reports. They are highlighting which activities work well and which do not, strengthening the overall commitment and how they impact the overall goal of well-being.

This 'design with the end in mind at the outset of design' methodology should already be a part of your Responsible AI design practices to avoid unintended consequences of 'ethics as an afterthought.'

WHAT TYPE OF ORGANIZATION (CORPORATE, NONPROFIT, NGO) ARE YOU?

Beyond readiness of any AI organization for the introduction or utilization of well-being metrics, the structure of an organization may affect how quickly these new metrics are adopted. Some initial (general) points of consideration to include during assessment:

- **Start-up:** The traditional venture capital model makes it difficult for start-ups to prioritize well-being metrics if they are working to generate the largest initial public offering (IPO) possible in the fastest amount of time. This legal and financial structure is one of the most difficult to address in terms of incorporating long-term, sustainability-oriented well-being metrics. Fortunately, organizations now stress that companies utilizing increased ESG reporting in a holistic, operational way make better long-term investments. This trend could lead to a different approach to acquiring start-ups that can and should prioritize well-being metrics for AI.

- **Enterprise:** Larger for-profit companies may have the greatest ability to adopt holistic well-being metrics due to existing employee roles/divisions such as CSR. They also hold the most power and resources to impact change. End-of-year reporting that includes environmental impact or societal issues is becoming the norm. Advocates for Responsible AI can go to department heads dealing with CSR and ESG issues and ensure that reporting includes front-end Responsible AI KPIs.

- **Not-for-profit/nonprofit:** Depending on size, nonprofits, especially ones utilizing progressive legal structures such as B-Corporation models, tend to lead the way in aligning well-being metrics with their brands or missions.

There is a difference between aspiration and application. Organizations can desire to change their reporting structures. Still, we recognize that an organization's operations are shaped by maximizing profits and growth as a primary KPI, not just their marketing messages or what could be obligatory and after-the-fact reporting.

WHAT IS YOUR ROLE IN YOUR ORGANIZATION?

Different people within an organization will have different starting points. For example, you may already be an ethics professional who has a mandate to change the design of your products to promote values like well-being. You may be an executive who can push for a department or a role to be set up to take accountability for this subject and be given the resources to succeed. You also have power in a middle management or junior position within an organization. In this case, it is wise to spend some time campaigning for the time, money, and resources to make a difference. For more on this point, please see Hattusia's report on "Net worthy: How to get budget and buy-in for tech ethics (and other forms of responsible business), "gaining budget and buy-in for technology ethics and other forms of responsible business.

The evolution of legal and regulatory frameworks on ESG reporting is a crucial way well-being metrics can and will evolve in the future. The existence of regulations regarding AI provides a pivotal opportunity to grow from a mindset of compliance to one of innovation in the face of mandated governance. For instance, when the European Union's (EU) General Data Protection Regulation (GDPR) first came into existence, many people fought against having to use it. But when GDPR was introduced to employees, there was a new clarity and understanding of personal data. The work concerning compliance brought clarity. It is highly likely that well-being, and related indicators, will have a similar uptake.

DEFINING THE METRICS OF RESPONSIBLE AI—MOVING FROM WELLNESS TO WELL-BEING

When we speak about "wellness," we are referring to the typical framing of "employee wellness programs" as defined by finite or discrete portions of a person's physical or mental health as aided by a Human Resources department. These efforts provide a fundamental basis for a long-term, holistic, and ongoing flourishing for all people supported and mirrored within economic and societal metrics we refer to as "well-being" in this paper.

Well-being metrics—which include ESG, SDGs, or any global metrics measuring areas beyond fiscal attributes—have a direct impact on companies and their stakeholders as well as on people and the planet beyond their walls.

There are dozens of these metrics, but the UN SDGs are the most well-known well-being indicators. The UN has stated that all 17 SDGs must be fulfilled by 2030 for societal success and sustainability. Their holistic application provides perspective for how to make paradigm changes, prioritizing issues like eradicating hunger.

The work of the WEF's International Business Council (IBC)—featured in the report, "Measuring Stakeholder Capitalism: Towards Common Metrics and Consistent Reporting of Sustainable Value Creation"—provides a set of 21 core and 34 expanded metrics for organizations to consider in the mix of reporting. Utilizing these metrics as a basis for Responsible AI design is an early recommendation to help assess your organization's preparedness in the well-being metrics space overall. While this is not AI-specific, it provides high-level logic for grouping metrics and risk in a way that can address regulation regarding digital/technological issues and those related to sustainability and ESG/CSR reporting.

A key point to start with in conversations about these issues is to use a tool like the WEF metrics to see how they honor existing organizational values. Using metrics like these as a conversation starter can help evolve a company's culture—and eventually its operations or business model—when framed as opportunities for *responsible innovation* writ large versus focusing solely on areas of artificial intelligence (meaning one technology in isolation) or potential regulation.

The good news is that the tide is turning regarding well-being (ESG/SDG) reporting. For instance, shareholders are now calling on their companies to provide richer environmental, social, and governance (ESG) reporting focused on long-term sustainability, as our planet needs restoration (not just avoidance of further harm). Companies are now also tying executive pay to ESG performance.

Utilizing well-being indicators and tools like the Well-Being Impact Assessment (WIPA) process in IEEE Std 7010™-2020 and other sustainability tools means we can use well-being indicators to guarantee people and the planet are being served by our finances versus the other way around.

17 SDGs

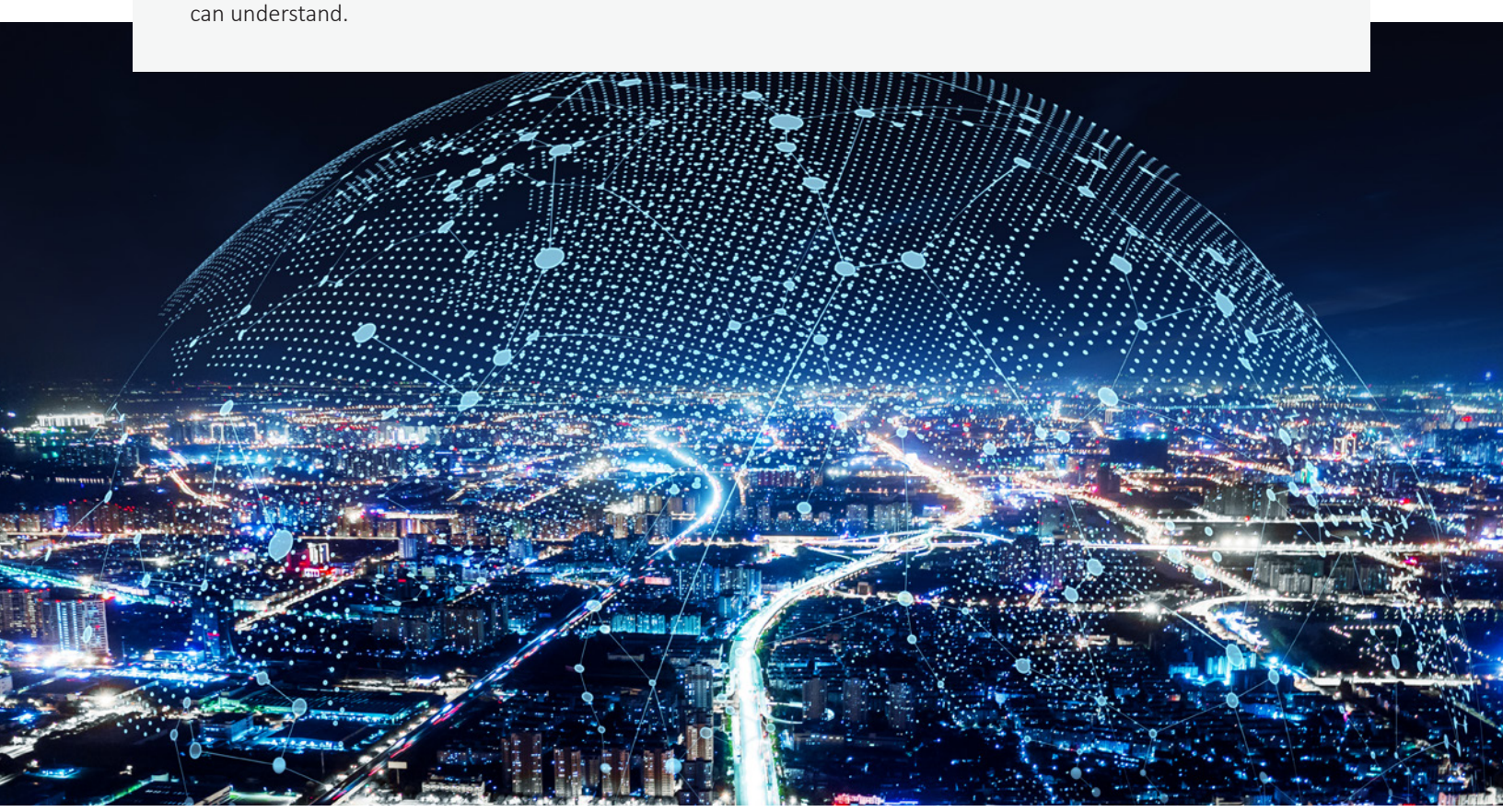
The UN has stated that all 17 (SDGs) must be fulfilled by 2030 for societal success and sustainability.

OUR RECOMMENDATIONS TO ADOPT THE VALUES NEEDED FOR RESPONSIBLE AI

Identify inhibitors. Adopting and aligning on well-being values may not always be easy. Some may fail to see the benefit of values or principles until these concrete practices are operationalized. Others might believe well-being is CSR teams' purview, not strategic priorities. Some may point to limited resources (lack of funding or teams spread across multiple initiatives), making this work easiest to deprioritize. Will the organization's culture inhibit the adoption of well-being as a strategic priority? This skepticism reinforces the need for well-being to be a strategic priority supported by leaders for these initiatives to take root and succeed.

SOME PRACTICAL ACTIONS TO ADDRESS THESE INHIBITORS ARE AS FOLLOWS:

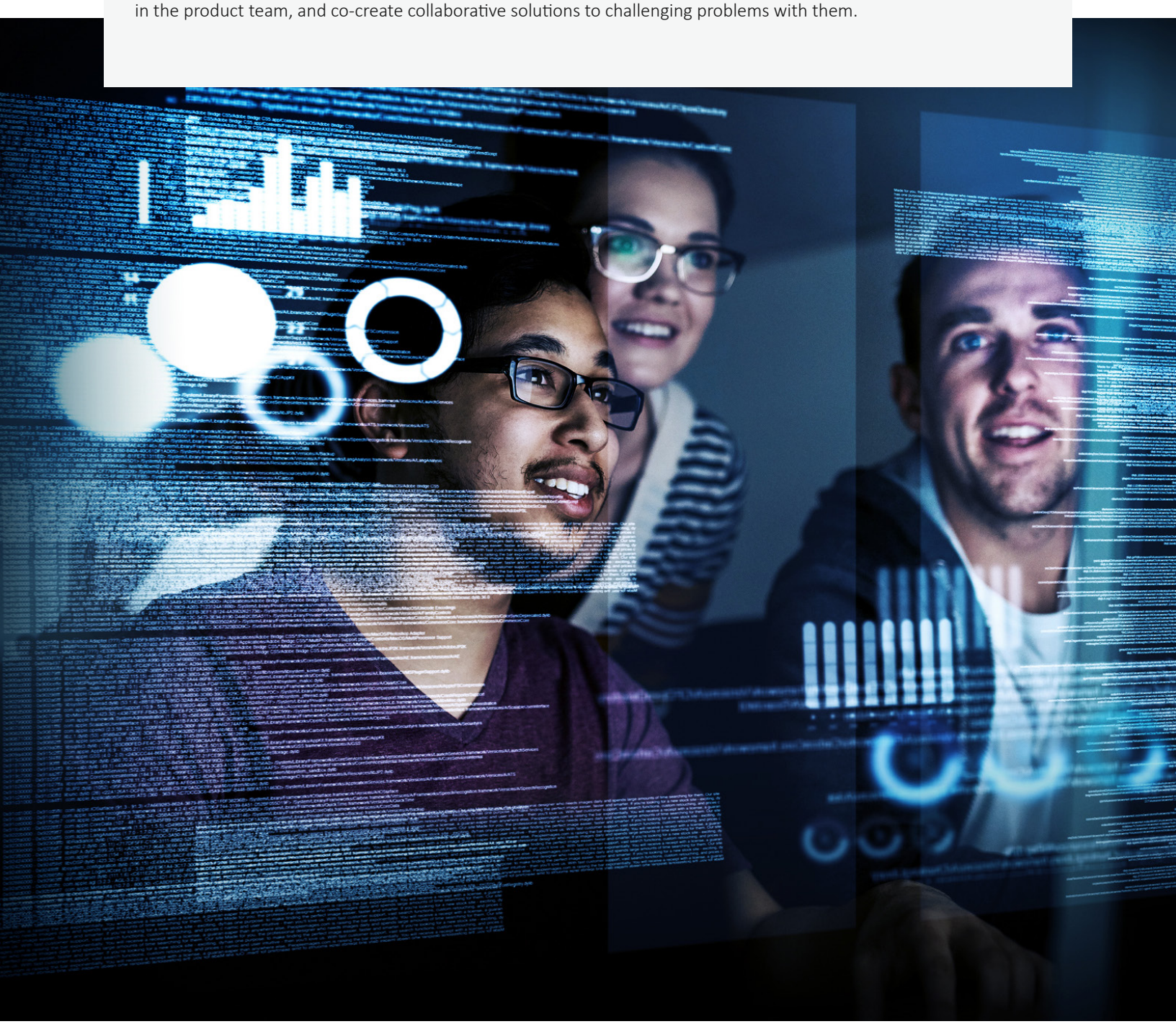
- **Consequence scanning workshops.** Doteveryone's manual offers an introduction to the overall logic of AI ethics/responsible innovation, which is a way for organizations to consider the potential consequences of their product or service on people, communities, and the planet. This innovation tool also provides an opportunity to mitigate or address potential harms or disasters before they happen to help ensure that the products or services they are creating align with their organization's values and culture.
- **Assessing harm.** Certain companies have an excellent set of resources on harms modeling that puts legal issues like human rights in a context that technologists can understand.
- **Make the most of the metrics.** There is no such thing as "one well-being metric to rule them all." There are dozens of ESG reporting and CSR structures, which often vary by region and business vertical. It is essential to understand that metrics may conflict with one another—or seem to—based on what you are trying to measure and what you wish to prioritize. Using societal impact assessments like the one on well-being from IEEE Std 7010™-2020, IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems can be helpful, as well as taking the B Impact Assessment.

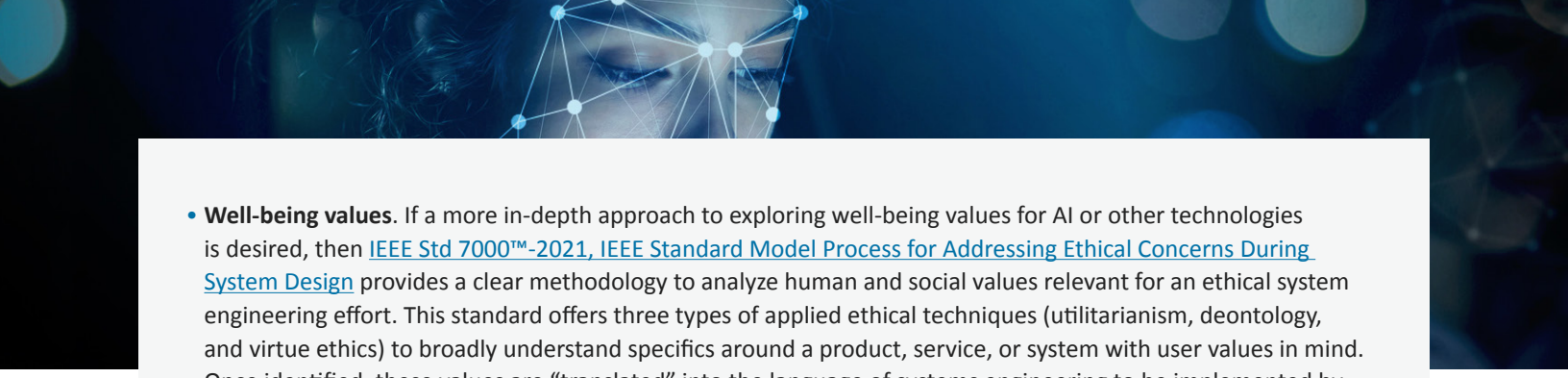


Educate and recruit. One of the most significant assets a company can leverage is its own people. Staff members are likely looking for ways to contribute to the overall organization's strategy, so leverage their initiative to get them involved in your Responsible Innovation or AI ethics practices:

Community juries. Another engaging tool can be found in its community jury technique where:

diverse stakeholders impacted by a technology are provided an opportunity to learn about a project, deliberate together, and give feedback on use cases and product design. This technique allows project teams to understand the perceptions and concerns of impacted stakeholders for effective collaboration. A community jury is different from a focus group or market research; it allows the impacted stakeholders to hear directly from the subject matter experts in the product team, and co-create collaborative solutions to challenging problems with them.





- **Well-being values**. If a more in-depth approach to exploring well-being values for AI or other technologies is desired, then IEEE Std 7000™-2021, IEEE Standard Model Process for Addressing Ethical Concerns During System Design provides a clear methodology to analyze human and social values relevant for an ethical system engineering effort. This standard offers three types of applied ethical techniques (utilitarianism, deontology, and virtue ethics) to broadly understand specifics around a product, service, or system with user values in mind. Once identified, these values are "translated" into the language of systems engineering to be implemented by technologists who typically do not have formal ethical training.

- **Ask questions.** It is okay if it is tough to see the connection between employee wellness, responsible innovation, and AI. The technologies comprising AI (machine learning, deep learning, algorithms) can be extremely complex to understand at first. This is why asking questions or having more general discussions is a good place to start. Here are some samples to get you going:

  - How does our company view/address *wellness* for our employees? What are our metrics of success and how are they measured?
  - Does our company have any existing AI principles? If so, how do they mirror our values? Do they largely focus on internal behavior (codes of ethics for employees), or do they also address our customers, suppliers, and value chain? How are these principles measured (in public relations or marketing value or in other ways)?
  - Does our company use the term *well-being* in a broader way than *wellness*? Do we talk about the UN SDGs or other ESG metrics in regard to our AI products (either ones we use internally or that we produce for customers)?
  - Who focuses on wellness, well-being, or ESG metrics in our organization? Any or all of the following areas of a company may be involved; representatives from these departments should be invited to discussions:
    - Human Resources
    - Legal
    - Corporate Social Responsibility
    - Marketing (issues of corporate branding / how Responsible AI builds trust with customers
    - Sustainability
    - Accounting, where fiscal responsibility and metrics provide a roadmap for sociotechnical metrics
    - Supply chain management experts

We have provided some additional resources below to help get you started as you move from assessing these issues to applying more direct plans to integrate Responsible AI metrics in the next section.

Further Resources:

- Siokou C., R. Morgan, and A. Shiell, "Group model building: a participatory approach to understanding and acting on systems," *Public Health Research & Practice,* vol. 25, no. 1, 2014; e2511404.
- Dunn, Alix, "Kind Environments for technology organizations," The Relay, Apr. 28, 2021, https://relay.substack.com/p/kind-environments.
- Stray, Jonathan, "Aligning AI to Human Values means Picking the Right Metrics," Apr. 15, 2020, *Partnership on AI* (blog), https://medium.com/partnership-on-ai/aligning-ai-to-human-values-means-picking-the-right-metrics-855859e6f047.
- Partnership for AI "Framework for Promoting Workforce Well-being in the AI-Integrated Workplace," Aug. 27, 2020, https://partnershiponai.org/paper/workforce-wellbeing/.Hattusia, "Net worthy: How to get budget and buy-in for tech ethics," https://hattusia.com/net-worthy.

APPLY:

HOW DOES AN ORGANIZATION IMPLEMENT THE PRINCIPLES AND METRICS OF RESPONSIBLE AI?

"It took a full year to develop our trusted AI principles. I took them physically and digitally around the company beginning with engineers and managers eventually working my way up to all of our C-suite officers across the company. Every single one of those individuals signed off and understood what they meant for the company."

—Kathy Baxter, Principal Architect, Ethical AI Practice at Salesforce

This example from IEEE EAD for Business Committee member Kathy Baxter emphasizes the critical need of evangelists to champion Responsible AI and, by extension, any well-being metrics supporting their use and impact. Cultural transformation can begin with one or two people, especially when armed with data from key stakeholders such as shareholders, customers, or regulators. That way, when describing the need for Responsible AI or well-being metrics, evangelists can gain buy-in to applying insights from tactics and techniques previously described.

A key opportunity and mandate for all Responsible AI is risk management. Managing the unintended consequences of AI is still a nascent industry, which is why multiple organizations worldwide are working to create AI frameworks to address risk. The US Department of Commerce's National Institute of Standards and Technology (NIST) is working to develop a framework along these lines that, according to Lynne Parker, Director of the National AI Initiative Office in the White House Office of Science and Technology Policy, "will meet a major need in advancing trustworthy approaches to AI to serve all people in responsible, equitable, and beneficial ways. AI researchers and developers need and want to consider risks before, during, and after the development of AI technologies, and this framework will inform and guide their efforts."

The WEF expands on this opportunity in "A 5 -step guide to scale responsible AI" by recommending the need to build organizational capabilities.

Designing and deploying trustworthy AI systems should be an organization-wide effort. It requires sound planning, cross-functional and coordinated execution, employee training, and significant investment in resources to drive the adoption of responsible AI practices. To pilot these activities, companies should build an internal "Centre of AI Excellence," which would concentrate its efforts on two core functions: training and driving adoption. Indeed, to do their job, employees need to be trained to understand how risk manifests in their contextual interactions with AI systems and, more importantly, how to identify, report and mitigate them.

Along with the omnipresent threat of risks is the imperative to mention the growing value of AI. McKinsey & Company's global survey "The State of AI in 2021," featuring responses from 1843 respondents, points out that AI's prospects remain strong: "Nearly two-thirds say their companies' investments in AI will continue to increase over the next three years, similar to the results from the 2020 survey… The share of respondents reporting at least 5 percent of earnings before interest and taxes (EBIT) that's attributable to AI has increased year over year to 27 percent, up from 22 percent in the previous survey."

RECOMMENDATIONS TO IMPLEMENT THE PRINCIPLES AND METRICS OF RESPONSIBLE AI

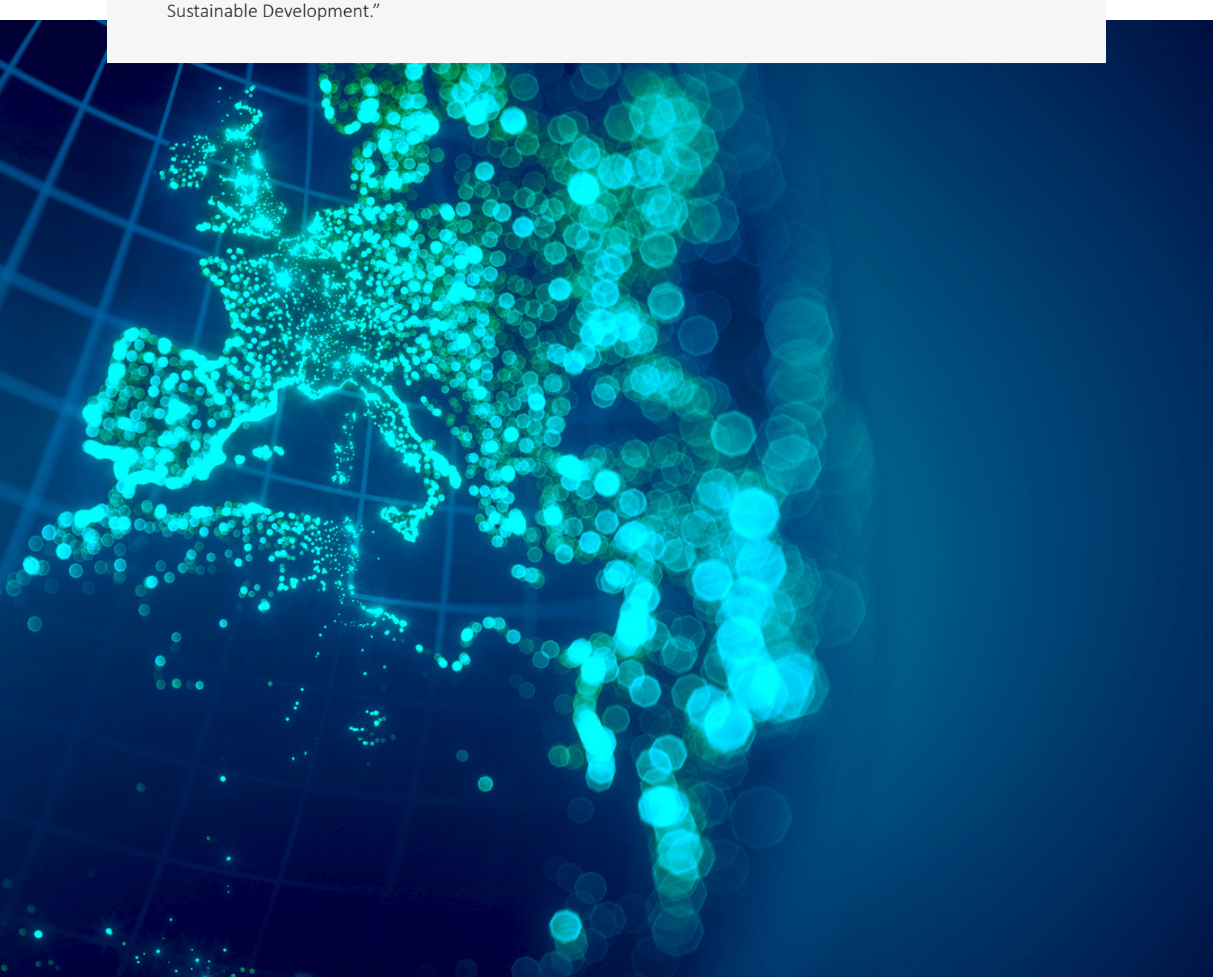
Suppose you are not using metrics to measure the effects of your product, service, or system on the planet or people. In that case, the risks and rewards you are measuring will primarily focus on KPIs relating to economics and growth. If your risk does not include how ecological systems could be damaged or human mental health may be at risk for your product, you guarantee "unintended consequences" in what you release. Note these could be positive and negative—so this is not a warning to just avoid regulation or be sustainable for sustainability's sake.

Start with your own role. This is a call for the due diligence required for those who want to champion and lead with genuine Responsible AI. Front-end design in this fashion is a form of research and development and innovation by definition. "How could our product increase SDG #5 focused on gender equality?" may seem like a difficult question, but not asking it can lead to those unintended consequences. When home-based vocal assistants first came on the market featuring female voices versus male voices, most people did not realize that children would start commanding their female teachers to do things based on their experience at home. These behaviors resulted in changing the design of AI-enabled voice assistants to feature male or gender-neutral voices. This is a perfect example of how utilizing SDG #5 in the design stage could have potentially avoided this issue.

Leverage AI to help you achieve your goals. There are many excellent resources to consider when thinking about the role of AI in achieving your agreed-upon metrics or the UN SDGs. Other business-oriented tools can help organizations use SDGs when designing their products, services, and systems to leverage AI. The following are some resources to help you move in the responsible direction:

- **"The role of artificial intelligence in achieving the Sustainable Development Goals."** This Nature Communications article written by numerous luminaries in the AI Ethics world, including Max Tegmark and Virginia Dignum, points out that:

  "…there is no published study systematically assessing the extent to which AI might impact all aspects of sustainable development—defined in this study as the 17 Sustainable Development Goals (SDGs) and 169 targets internationally agreed in the 2030 Agenda for Sustainable Development. This is a critical research gap, as we find that AI may influence the ability to meet all SDGs. Here we present and discuss implications of how AI can either enable or inhibit the delivery of all 17 goals and 169 targets recognized in the 2030 Agenda for Sustainable Development."

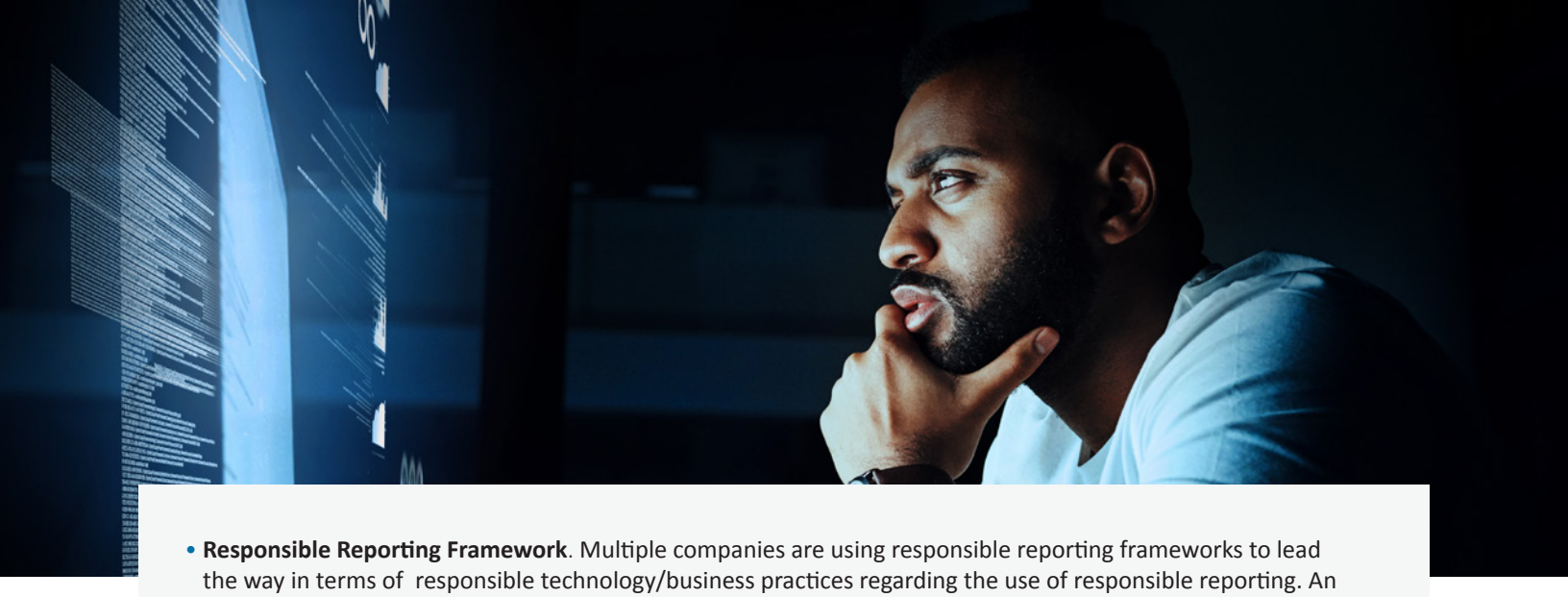


This extensive report provides an in-depth analysis of the areas where the development and use of artificial intelligence is either an enabler or inhibitor of each SDG, before finalizing the design of your AI, we recommend you refer to this report to see how your product, service, or system is either enabling or inhibiting specific SDGs.

- **The World Economic Forum.** Like the Nature Communications article mentioned above, WEF's article, "What Would it Take to Make AI Greener?" provides key insights into beneficial impacts that AI can have on our relationship to the environment as well. A comprehensive study in 2020 assessed the potential impact of AI on the United Nations' 17 Sustainable Development Goals, encompassing societal, economic, and environmental outcomes. The researchers found that AI could positively enable 93% of the environmental targets, including the creation of smart and low-carbon cities.

Tie to existing reporting frameworks and KPIs. The structures to measure progress for any Responsible AI metrics already exist in different forms. If your company is utilizing a specific reporting framework for KPIs, parallel the same structure and accountability for any Responsible AI metrics on which you have aligned. This will give your company a far more holistic picture of progress.

- **The SDG Compass.** Developed by GRI, the UN Global Compact, and the World Business Council for Sustainable Development (WBCSD), the SDG Compass incorporates feedback received through three consultation periods from companies, government agencies, academic institutions, and civil society organizations worldwide. The compass provides specific opportunities for businesses to use the SDGs as an overarching framework to shape, steer, communicate, and report their strategies, goals, and activities, allowing them to capitalize on a range of benefits. Specific examples (see p. 12 of the SDG Compass guide) also illustrate how you can map SDGs to the value chain of your business. The Inventory of Business Indicators is of particular note as it provides specific data-laden indicators you can use as KPIs regarding the design of your AI and the SDGs you are trying to focus on in your work.



- **Responsible Reporting Framework**. Multiple companies are using responsible reporting frameworks to lead the way in terms of responsible technology/business practices regarding the use of responsible reporting. An excerpt from their FY21 report shows the extent to which they use ESG and SDG reporting in as transparent a way as possible to guide their business.

  This report is also informed by leading ESG disclosure frameworks and standards, including the Sustainability Accounting Standards Board's (SASB) Software and IT Services sector guide, the Task Force on Climate-Related Financial Disclosures (TCFD), Global Reporting Initiative Standards (GRI), the Ten Principles of the UN Global Compact (UNGC), and the World Economic Forum—International Business Council's Stakeholder Capitalism Metrics (SCM).

  We incorporate the Sustainable Development Goals (SDGs) into our existing reporting processes to demonstrate our active participation as a business in advancing these goals. Transparency underlies Target 12.6 of the SDGs to encourage companies to adopt sustainable practices and integrate sustainability information into their reporting cycle. Through reporting we can better understand, communicate, and manage our contribution to the SDGs.

- **GSMA's AI Ethics Playbook for implementing the responsible use of AI.** Considering the diversities of different regions, the GSMA has published a practical guide to help organizations implement Responsible AI principles in day-to-day business. It also includes an online ethics questionnaire, allowing users to anonymously check the degree of ethical consideration of AI products and services. It is based on assessing the potential for harm (severity, scale, and probability), followed by a step-by-step evaluation of AI principles. The outcome is the measure of the ethics of the person who completes the questionnaire as guidance to better design AI and tech in ways that recognize, measure, and honor end user values.

Once you begin to measure and track well-being metrics through tools like the UN SDGs, it is time to discover how to amplify knowledge and successes across your organization and into its operations.

Further Resources:

- Wahl, Daniel Christian, "GDP as an insufficient economic indicator & some more systemic alternatives," Feb 17, 2017, https://designforsustainability.medium.com/gdp-as-an-insufficient-economic-indicator-some-more-systemic-alternatives-91d2e8c093df.
- Unilever "Sustainable Development Goals," https://www.unilever.com/planet-and-society/sustainability-reporting-centre/sustainable-development-goals/.
- Tetrapak, "Our Approach," https://www.tetrapak.com/sustainability/our-approach.
- IKEA, "IKEA Sustainability Report FY21," https://gbl-sc9u2-prd-cdn.azureedge.net/-/media/aboutikea/newsroom/publications/documents/ikea-sustainability-report-fy21.pdf?rev=6d09c40ec452441091b10d9212718192.
- United Nations, "Do you know all 17 SDGs?" https://sdgs.un.org/goals.

AMPLIFY:

HOW DOES AN ORGANIZATION SCALE RESPONSIBLE AI TO COMPANY CULTURE AND WIDESPREAD PRACTICES?

"We are accustomed to moving to answers, but we first have to understand the challenge and put it in context, in order to make the required changes in: Mindsets, Models, Structures, and Culture. So deepening our understanding of well-being and meeting the goals of SDG's, with effective use of AI, is the right starting point. Only then we can begin to rethink how our current systems are working, identify changes, and begin the process of internal co-creation of new systems and externally developing new partnerships that result in collaboration and new ecosystems."

—Deborah Hagar, President, Foundation for Sustainable Communities

Changing the status quo is always a challenge, mainly because there is comfort in regularity. We may not like multiple aspects of our job, life, or the world, but a status quo provides a sense of control. The good news about tying Responsible AI to human wellness or metrics on human well-being and ecological flourishing is that we can recognize where specific patterns that have formed a societal status quo aren't working for most people on the planet. By pointing a lens at Responsible AI metrics and the values underlying them, we can enable meaningful cultural shifts within and outside of our companies.

A crucial part of amplifying your work in Responsible AI must focus on looking for the elements most affecting your employees, partners, or customers that are not measured in meaningful ways today. There is a saying: "what you measure is what matters to you." But this is only partly true since people and their mental health matter whether or not society has prioritized those factors. The planet and its ecosystems matter whether we put protecting and regenerating Earth at the end of our fiscal or other priorities.

So if these things matter, should not a truly Responsible AI incorporate the planet and people into all design, manufacturing, and the entire life cycle of our products and services?

Like with all other ethical AI practices, well-being metrics are most impactful when they are thought out at the beginning of the design process—selecting well-being metrics/indicators at random risks misalignment and frustration. Companies should be aligned on the well-being metrics they are most suited to engage with and track at whatever size or maturity level. This enables an informed decision on what well-being metrics best suit the company and its mission.

WELL-BEING METRICS FOR LONGEVITY

Paying attention to employee flourishing and well-being can link to and improve organizational durability and long-term institutional flourishing. Organizations such as Cambridge University Press (founded in 1534), the French hospital Hotel-Dieu de Paris (founded in 651), and the Japanese construction company Kongo Gumi (founded in 578) are diverse in their activities and regional and cultural differences but share practices and values that indirectly promote sustainable well-being. Some of these practices and values are as follows:

- The products and processes related to these organizations are explicitly connected to timeless public purposes that serve a common good in their corresponding fields.
- The identity of these organizations is blended with that of specific communities (of workers as well as consumers) who nurture virtuous values.
- The activities of these organizations are developed by skilled and experienced professionals (printers and authors in the case of Cambridge University Press and engineers in the case of Kongo Gumi) who are highly valued by and retained in the organizations.
- The products of these organizations directly relate to either basic needs (such as healthcare and housing) or transcendental values (such as knowledge and religion).

Assuming that longevity is a consequence of nurturing such practices and values, by incentivizing longevity, we can encourage organizations toward the holistic promotion of well-being.

Our Recommendations to Scale the Success of Responsible AI and Well-Being Metrics

The following recommendations can further incentivize an organization to scale well-being metrics as success criteria for Responsible AI:

ADJUST GOVERNANCE FOR YOUR SIZE OF OPERATIONS AND YOUR AUDIENCE

- **Start-up:** Smaller companies may see governance and processes as barriers. They do not have the same resources as large enterprises, so allocating even one person full-time to manage the implementation of these metrics is a significant investment that cannot always be met. Therefore, a bottom-up cultural approach is better suited to smaller companies. Instead of constant monitoring, the metrics are instilled into company culture. As the company grows and matures, these natural habits are slowly translated into processes.
- **Enterprise:** The larger the company, the more top-down governance is required to execute successful adoption. Larger companies need cross-departmental guides; otherwise, efforts become siloed or lost in overall translation. Additionally, it is easier to communicate/gain consensus on values and well-being in smaller groups. The more people involved, the easier it is for individuals to get lost in understanding and implementation.
- **Non-profits:** Non-profits tend to scale through external partnerships and benefit from being value-driven. Depending on their size, they could combine both a startup and an enterprise approach to best showcase and differentiate themselves by those very values.

Leverage existing roles. Various roles and duties can be leveraged to operationalize Well-being by Design. Establish clear stakeholders/owners of the metrics with individuals that will help drive the initiative forward and take responsibility. Without such champions, well-being can get pushed aside as a "nice to have, but we are too busy right now."

No matter the size of the organization, leadership buy-in and culture matter. Suppose there is not sincere intention or motivation within the company to adopt these metrics. In that case, they will only become a burdensome checkbox that frustrates employees and negates any real benefit that the metrics were meant to bring in the first place. For the operational implementation of relevant roles, much can be leveraged (i.e., learned and reused) from the data protection officer (DPO) function, which many organizations have in place. However, ethical AI also introduces new aspects that may require an additional role. Certain companies have identified a "Responsible AI Champion" (RAI Champion) who is the go-to person in case teams have questions or doubts about ethical aspects of using AI. An RAI Champion would have the following responsibilities: inform, educate, advise and escalate, coordinate, connect, and manage change.

Consider a safety culture. Prioritize the wellness of employees, mirroring a focus on well-being/Responsible AI metrics for what you offer your customers and stakeholders. You can read a good overview of this logic about Hudson's Safety Ladder here or check out Montana State University, which describes its Positive Culture Framework as follows:

We believe the positive exists and is worth growing. We reject using fear to motivate change. The Positive Culture Framework focuses on revealing and growing positive, shared values and beliefs, which result in healthier and safer behaviors. This perspective does not mean that we ignore or minimize the negative, harm, or risk that communities face. The Positive Culture Framework raises both concern and hope to foster engagement, create energy, and bolster self-efficacy to move forward.

Agile marketing for sustainability. "Leveraging the Agile Manifesto for More Sustainability" from InfoQ discusses how "the Agile Manifesto can be leveraged for increasing sustainability. For example, when focusing on businesspeople and developers working together, explore the target group and ensure that you embrace accessibility." This resource can be a way to introduce ideas of well-being metrics and Responsible AI via agile marketing procedures you may already have in place (e.g., scrums).

Ethical conformity assessment. This instrument, such as IEEE's CertifAIEd™ framework, affirms an organization's commitment to upholding human values, dignity, and well-being and to respecting, protecting, and preserving fundamental human rights. Certification guidance and assessment as well as independent verification offers the ability to scale responsible innovation implementations, thereby helping to increase the quality of AI systems, the associated trust with key stakeholders, and realizing associated benefits. Conformity assessment will be mandatory for high risk applications in the EU.

The AI Ethics maturity model. Written by Kathy Baxter of Salesforce (who is also a committee member), this model provides an excellent and thorough tool from the same company that is leveraging UN SDGs and ESGs for sustainability (as mentioned above).

Further Resources:

- Partnership for AI "Framework for Promoting Workforce Well-being in the AI-Integrated Workplace," Aug. 27, 2020, https://partnershiponai.org/paper/workforce-wellbeing/.
- Lee, Allison Herren, "A Climate for Change: Meeting Investor Demand for Climate and ESG Information at the SEC," (Speech), U.S. Securities and Exchange Commission, https://www.sec.gov/news/speech/lee-climate-change
- Edelman, "Edelman Trust Barometer 2021," https://www.edelman.com/sites/g/files/aatuss191/files/2021-03/2021%20Edelman%20Trust%20Barometer.pdf.

CONCLUSION:

WELL-BEING BY DESIGN (WBbD)—THE NEW DEFINITION OF RESPONSIBLE INNOVATION IN AI

As a general rule, success in designing products, services, and systems traditionally relies on customers or end users buying or using whatever appeals to them. In pre-internet days, what appealed to people may have come via word-of-mouth from friends, radio, or television advertisements.

Today, direct messaging targeting people to purchase goods and services based on their location, identity, and data can be delivered dozens of times a day via invisible algorithms. It is also created in ways that are not transparent to users or that go against their wishes. The success metrics for these algorithms are not based primarily on increasing a customer's long-term flourishing (a state of overall, continuous physical and mental health) or purpose but on data monetization and attention hijacking. This is the KPI of success: Did the work that went into reaching this specific individual motivate them to click a button and spend money as a proxy for their short-term need or happiness? Or did it divert and hijack their attention as a way to monetize?

There is a lack of connection between the trillion-dollar ad industry and the in-depth connection to emotions, mental health, and worth to the idea that we "can not measure well-being." Many times when you use the term *well-being*, people understandably think we are talking about short-term mood or one person's happiness; they think these things can not be measured. Yet, these are precisely the things the advertising industry does with the highest levels of AI and other technologies available today.

Perhaps we don't use these same tools to measure long-term well-being because we have not recognized our human need for long-term flourishing. And in terms of the corporate world, the concept of *Well-being by Design* (provided as a phrase encompassing the overall recommendations of this paper) requires not only a recognition of changing business models for long-term sustainability but a change in how all AI products, services, and systems are created in the first place. Responsible innovation is not "responsible" as long as "societal success" continues to be defined by the short-term mandate to rush to market and the deprioritization of people and planet.

WELCOME TO THE RESPONSIBLE AI REVOLUTION.

WELL-BEING METRICS READINESS FRAMEWORK[6]

	Lagging	Basic	Advanced	Leading (WBbD*)
Internal training, support, and people resources *(Who or what exists at an organization to introduce, support, or drive well-being metrics)*	• Employees on their own to use any metrics beyond standard financial KPIs • May be encouraged but no official support	• Well-being efforts come from outside consultants and focus largely on employee health	• ESG, CSR reporting considered but not prioritized at the outset of design • Training for financial or other reporting includes recognition of design-oriented ESG or well-being metrics	• Well-being metrics are understood and utilized by all employees (roles) and in onboarding
Leadership buy-in	• Leadership is unaware of or won't implement well-being metrics	• Introductory workshops or trainings on well-being metrics provided	• Leadership explores proof of concept tests around product or service design utilizing well-being ethics methodologies	• Leadership prioritizes and requires well-being metrics as the top KPIs for design, value, and company brand
Metrics and KPIs	• Organization only utilizes traditional single bottom-line KPIs	• ESG, CSR reporting created as an afterthought or solely for reasons of compliance	• Well-being efforts include the trial of methodologies such as impact assessments and process or governance models around holistic well-being metrics	• Comprehensive and holistic well-being metrics are utilized as KPIs for design at the outset of ideation • Well-being metrics are weighted alongside financial KPIs
Organizational impact	• Well-being metrics are not understood or purposefully ignored • Any existing efforts are siloed	• Well-being metrics begin to resonate through all operations (not just CSR) • Longer-term well-being goals are set for the company	• Well-being metrics are formally adopted as reporting beyond standard ESG or CSR reporting	• Well-being metrics and their use define the brand messaging • Well-being metrics increase employee retention

**Well-being by Design—The new standard for responsible innovation*

[6] This framework mirrors the AI Ethics Readiness Framework that is featured on page 13 of the first document created by this committee, "A Call to Action for Businesses Using AI". We recommend comparing these to frameworks if you work in AI as a way to incorporate well-being metrics into your work.

COMMITTEE MEMBERS:

Chairs:

- Adam Cutler — Distinguished Designer, AI Design, IBM
- Milena Pribic — Senior Designer, AI Design, IBM

Committee Members:

- Uthman Ali — Digital Ethics Lead, BP
- Kathy Baxter — Principal Architect, Ethical AI, Salesforce
- Richard Benjamins — Chief AI & Data Strategist, Telefónica; Co-founder of the Observatory for Social and Ethical Impact of Artificial Intelligence
- Flavio S. Correa da Silva — Associate Professor, University of São Paulo.
- Olivia Gambelin — Founder, Ethical Intelligence
- Ilana Golbin — Responsible AI Lead, PwC
- John C. Havens — E.D., The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems & Sustainability Practice Lead, IEEE Standards Association
- Kashyap Kompella — FHCA, Founder, AI Profs; CEO, RPA2AI Research
- Carl-Maria Mörch — Manager FARI - AI Institute for the Common Good (ULB-VUB)
- Clara Neppel — Senior Director, European Operations
- Alice Thwaite — Founder, Hattusia

IEEE SA STANDARDS ASSOCIATION

Visit us at **standards.ieee.org** and **connect** with us on:

- **Facebook:** facebook.com/ieeesa
- **Twitter:** @ieeesa
- **Linkedin:** linkedin.com/company/ieee-sa-ieee-standards-association
- **YouTube:** youtube.com/ieeesa
- **Instagram:** instagram.com/ieeesa
- **Beyond Standards blog:** beyondstandards.ieee.org

IEEE