DO NO HARM

How GPs and LPs can use Responsible AI to build trust

JULY 2024

STEPSTONE

stepstonegroup.com

Table of contents

Introduction

The internet changed everything. Although few actually understood how it works, nearly everyone understood that it would revolutionize life. Fewer still contemplated the negative side effects it would one day pose. Without much in the way of restraint, the internet's growth was swift and unbridled. As the modern world embraces another game-changing technology, armed with the benefit of hindsight and years of study on the internet's harmful side effects, society by and large is much more cautious about the development and application of artificial intelligence (AI)—especially generative AI.

Though it is still in its infancy, we seem to have a much greater understanding of the risks inherent in generative AI than we had when the internet was a similarly fledgling technology. Our collective anxieties about where an adaptive and autonomous technology might lead us have steered us to the development of the emerging field of "Responsible AI." Because most of the existing processes and tools, from code development to risk management, were designed for traditional software systems, they struggle when dealing with generative AI systems, being ineffective at managing emergent risks and preventing harmful outcomes. Responsible AI will be critical in responding to such challenges and delivering trustworthy AI systems.

Because GPs and LPs are involved in both developing and applying AI to the companies and assets they invest in, there is a vested interest in ensuring that AI is developed and deployed responsibly. We recognize the immense benefits these systems could deliver—both financial and otherwise. And while there is huge collective focus on this upside, to ensure this materializes Responsible AI practices need to develop. Today, this is a nascent field. Our firm, through this paper and related efforts, hopes to contribute to its development. Leveraging existing ESG frameworks and expertise in value creation will be helpful to this end.

This paper provides an overview of the scope, history, global initiatives and regulatory developments of AI broadly encompassing generative AI. It introduces the concept of Responsible AI, which seeks to deliver trustworthy AI systems. The risks endemic to these systems are explored as are the newest leading AI risk-management frameworks. This paper seeks to contribute to the nascent practices of Responsible AI in private markets, providing examples and suggested best practices at the GP, LP and asset levels. We explore how ESG practices dovetail with Responsible AI and pay close attention to "high-risk sectors."

Defining AI

AI is commonly thought of as a technology that enables computers and machines to simulate human intelligence and problem-solving capabilities. This definition isn't wrong per se, but it is lacking. OECD AI, a leading think tank, offers the following definition, which was revised in 2023 to encompass generative AI considerations like autonomy and adaptiveness:

> *"An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment."*

This definition introduces some important concepts that lie at the heart of Responsible AI—namely, how AI affects humans, and influences physical and virtual spaces.

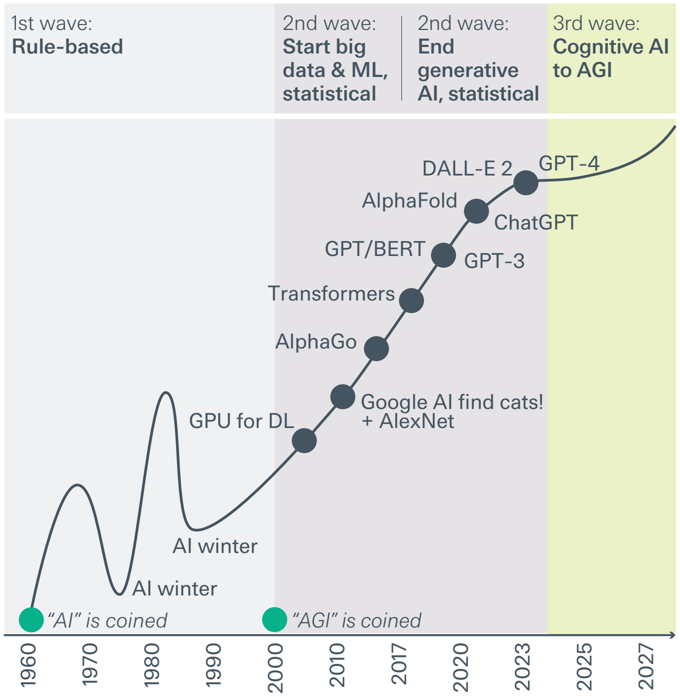
A brief history

Since AI was envisioned in the 1950s it has enjoyed waves of development interspersed with winters of stalled progress. Today we live in the era of "narrow AI" in which AI can perform discrete tasks like creating personalized playlists and news feeds or helping an Uber driver find the most efficient route.

Traditionally, AI was embedded in programs and layered into applications by specialist programmers. With the arrival of ChatGPT, the doors opened for laypeople to access and apply generative AI, a specific subset of AI models that can generate content (in all forms—text, audio, visual) enabled by large language models (LLMs).[1] With this accessibility, there was an explosion in usage.

Even though the applications are becoming more complex, we have a way to go before we see "strong AI." Also known as artificial general intelligence (AGI), strong AI refers to AI systems that possess humanlike cognitive abilities. Arguably, this is what most think of (and fear) when they think of AI.

FIGURE 1: TIMELINE OF AI SYSTEMS



Source: Voss, Peter & Jovanovic, Mladjan. (2023). "Concepts is All You Need: A More Direct Path to AGI."

[1] Such machine learning systems have in turn been enabled by greater computing power and availability of large datasets (courtesy of data scraping).

A dynamic landscape

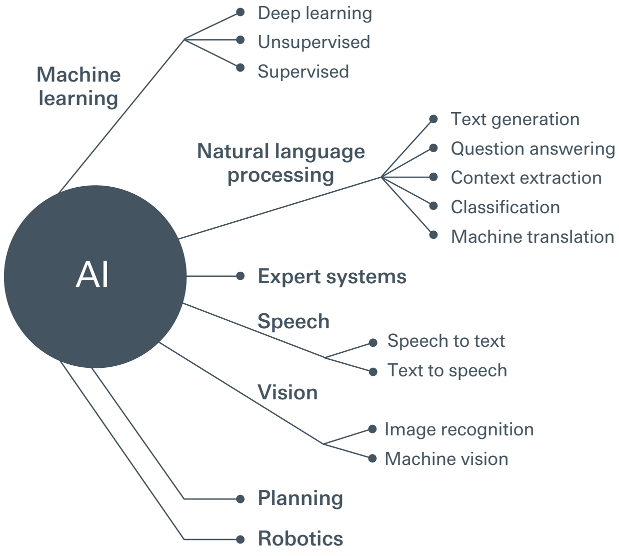
Most experts think of AI in one of two ways: either as a shifting landscape made up of several elemental applications (**Figure 2**) or as a life cycle system (**Figure 3**). Each mode of thinking is instructive in its own way.

- The "landscape view" illustrates AI's complexity and the challenges associated with structuring a framework to govern it.
- The "system view" considers everything from design and collecting the data that will inform the AI to verifying and validating the system. Most of the think tanks and lawmakers contemplating Responsible AI adhere to this view.

The system view articulates three things that are important in the context of Responsible AI:

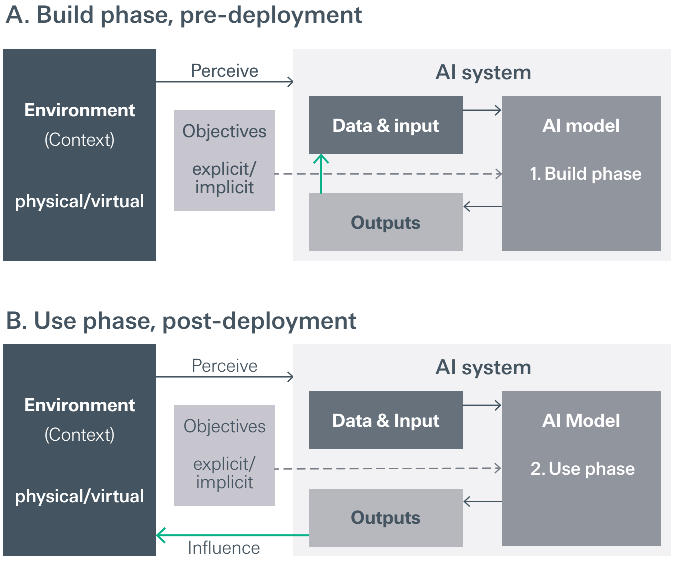
1. AI systems can exert influence on the environment—physical or virtual;
2. This influence can be positive or negative; and
3. Owing to AI's inherent autonomy (most relevant for generative AI), that influence is not explicitly defined or controlled by humans.

FIGURE 2: THE LANDSCAPE VIEW OF AI



Source: StepStone Group analysis.

FIGURE 3: THE (SIMPLIFIED) SYSTEM VIEW OF AI



Source: OECD, 2022.
Note: This figure presents only one possible relationship between the development and deployment phases. In many cases the design and training of the system may continue in downstream uses. For example, deployers of AI systems may fine-tune or continuously train models during operations, which can significantly affect the system's performance and behavior.

AI, robot

Sci-fi is replete with examples that stoke our fears of automated vehicles striking pedestrians or an all-intelligent supercomputer wreaking havoc on the financial system.

Based on the writings of Isaac Asimov, the classic film “I, Robot” is one such example. In it, a form of strong AI attempts to crush humankind by corrupting and co-opting an army of robots, which have hitherto been governed by Asimov’s “Three Laws of Robotics.”

An early form of Responsible AI, these Laws, which seek to ensure that robots cannot harm humans, underscore the importance of thoughtful technological advancement. That they alone are not enough to put a stop to the strong AI (Will Smith helps save the day) illustrates the challenge that technologists, luminaries and lawmakers face in developing a framework for Responsible AI. Frameworks like Asimov’s Three Laws are critical, but human engagement is too.

The three laws of robotics

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Introducing Responsible AI

Much of our collective understanding of how technology behaves is rooted in the fact that humans are in control of the tech's design, production and employment. This is true of everything from computer programs to automobiles. And while each of us may drive differently, we all operate our cars in pretty much the same way. In such linear systems, where outcomes are repeatable and nearly certain, past results are almost always prologue. This mode of thinking, however, is inadequate to understanding newer forms of AI, such as generative AI.

AI learns from a vast underlying dataset and has some autonomy in interpreting it. As anyone who has used DALL-E can attest, the results aren't always repeatable. More problematic still is that the data might be interpreted in a vacuum—or worse, incorrectly—thus opening the door for AI to reinforce biases and prejudices. "Algorithmic redlining" has entered the popular vernacular.[2] In summary, the scale and complexity of AI complicates our traditional approach to designing, developing and testing tech. This is even more acute with generative AI.

As such, a worldwide effort is underway to establish best practices for building and managing Responsible AI systems that meet the guidelines set by the US's National Institute of Standards and Technology (NIST), as shown in **Figure 4**.[3]

Grappling with trade-offs

At its core, Responsible AI is about grappling with trade-offs.

- A secure system may be opaque and hard to interpret—part of what makes it secure.
- A reliable system that produces valid outputs may be opaque with low explainability.
- The data undergirding one system might be profoundly biased, and yet that system could produce results that are valid yet hard to interpret.

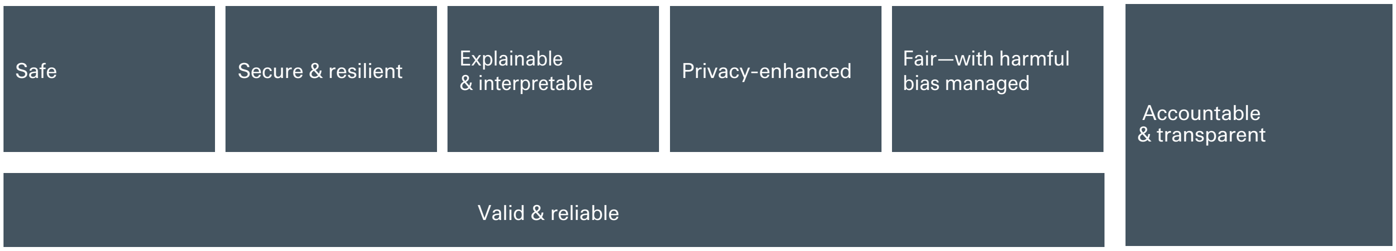
This means that all stakeholders affected by AI systems, from developers to users, are explicitly or naively trading one thing for another. Since there is no AI system without trade-offs, the question becomes: What is needed to make an informed decision? How can we know whether a system meets a

FIGURE 5: RESPONSIBLE AI TRADE-OFFS

Threshold for safety	Value
Data quality (bias, discrimination, privacy, safety)	Low ←◆——————→ High
Data relevance (relevant/disconnected)	Disconnected ←———◆→ Relevant
Model objectives	Explicit ←——◆—→ Implicit in data
Explainability (auditability/repeatability)	Low ←◆——————→ High

Source: StepStone Group analysis. For illustrative purposes only.

FIGURE 4: CHARACTERISTICS OF TRUSTWORTHY AI SYSTEMS


Safe
Secure & resilient
Explainable & interpretable
Privacy-enhanced
Fair—with harmful bias managed
Accountable & transparent
Valid & reliable

Source: NIST AI Risk Management Framework, 2023.

[2] In other words, being overly reliant on algorithms to make sensitive decisions about loans or hiring can be unfairly or illegally discriminatory.
[3] Despite its naming convention and the fact that many of its concepts nest neatly into responsible investment and ESG, Responsible AI was not born of either. Still, asset managers and owners with strong ESG integration practices may be better able to address AI challenges.

Responsible vs. Ethical AI

Within the literature broadly, Responsible AI and Ethical AI are often used interchangeably. Though some might argue the distinction is an important one, we think the point is altogether moot. Even a cursory comparison of the OECD's AI principles and UNESCO's Ethical AI principles shows a lot of overlap and only slight differences. In the march toward Responsible AI, the destination may very well be more important than the path taken to get there.

Ethical AI seeks to ensure that AI systems align with a particular value set (e.g., reinforcing democratic values) in addition to embedding a range of normative values such as transparency or absence of bias. Responsible AI, on the other hand, focuses on normative principles, which we find more conducive to broad adoption irrespective of stakeholder jurisdiction.

We have adopted the nomenclature of Responsible AI because even though not all organizations or nations will align on ethical positioning, they can still support the responsible development and deployment of AI.

minimum threshold for safety, validity or explainability? Right now we can't. One day we might see "CE" or "ISO" stamps of approval that signal an AI system meets such minimum thresholds.

Key issues—risks

Responsible AI frameworks are emerging in the wake of the growing awareness of the risks posed by AI. These risks might lead to harm, which could threaten specific rights.

Importantly, these frameworks have been a critical precursor to the emergence of regulations globally.[4] **Figure 6** explores the key risks endemic to generative AI.

As these risks increase, so too does the probability that harm will come to people (e.g., safety), organizations (e.g., business operations) and ecosystems (e.g., the global financial system).

If the risks and resultant harms are considered through a sector lens—particularly the sector where the end user resides—then it becomes apparent that risks are likely to be more prevalent in

FIGURE 6: KEY RISKS IN GENERATIVE AI

Discrimination/ bias/fairness Data sets driving and reinforcing discriminatory/biased outcomes	LLMs require massive datasets that are often built via bot scraping. It is not uncommon for multiple datasets to be combined. Determining the embedded bias is tricky and it often emerges only in application outcomes. Even then different biases may exhibit inconsistently.
	Examples • Recidivism algorithm predicts black people twice as likely to reoffend • Search algorithm exhibits gender bias • Debt approval system biased on race and gender grounds • Image generator reinforces gender and racial stereotypes (e.g., an "assistant" is a young woman; an "executive" is an older white male)
Data protection IP protection/data approval/consent	• Data scraping leads to a range of IP/consent issues • Facial recognition data viewed as higher risk with higher consent requirements • Continuous surveillance at home, work, and in education and health environments viewed as invasive and threatening freedom of association • Deepfakes and fake news drive harm, threaten freedoms and systems of government
	Examples • 2023 Federal Trade Commission investigating OpenAI with respect to consumer law breaches. OpenAI response with GPTBot (web crawler), which allows easier blocking of its data scraping. • In USA, current legal cases on applicability of concepts of fair use versus copyright in relation to training datasets. If this is deemed fair use, developers could use copyright data in training sets. • Can AI generate copyright or patented outputs? Today's response varies by jurisdiction. • AI coding assistants enabling automated cybersecurity attacks • AI's role in mass surveillance and censorship

[4] See page 20 for more information.

FIGURE 6: KEY RISKS IN GENERATIVE AI (CONT.)

Explainability Understanding why model output is what it is	• Outcome of model aligned to model objective; and outcomes exhibit consistent results (i.e., repeatability) • Data appropriately contextualized • Auditability enabled by appropriate disclosure and model transparency
	Examples • Amazon's CV review system penalized women; it took months to understand why • Geographic biases in training datasets can result in racial bias exhibited in predictive policing software • Image generators reinforce gender and racial stereotypes (e.g., an "expert" is an older white male) • Certain medical diagnoses show less applicability for black population because of bias in training set
Human control/ appeal Choice for a human route	Ability to appeal to a human, particularly in high-risk situations that can affect human life or safety
	Examples • Overriding a drone attack • Appealing a rejection decision generated by loan software • Deviating from set transportation/logistics schedules • Having the right to appeal automated legal decisions
Human centered/ respect for human values System is designed for the benefit of humans individually and collectively in society	• The objective of the model is to not result in outcomes that harm humans • Human harms include physical, psychological, denial of access to work/finance/education/ housing/human-centered values including democratic rights/self-expression/safety from hate
	Examples • Content moderation driving social echo chambers • Concept of collective disempowerment—model takes on increasingly important functions in society • Concept of power seeking—AI designed to drive power accumulation • Above issues potentially exacerbated by concentration of power in limited MNC developing AI capability

Source: StepStone Group analysis.

FIGURE 7: RISING RISKS, RISING HARMS

Increase in risks/harms

Defense/law enforcement	• Encompasses weapon systems, policing systems, etc. • Extensive concern about harm to humans resulting from decisions • For some systems, determining how to allow for human intervention, while critical, is complex • Extensive data privacy/data security concerns • High risk for bias in datasets leading to discriminatory outcomes • Use of facial recognition for policing and defense purposes generally allowed under regulation globally—raises the challenge of balancing rights of individual versus security of society	Examples include drones, weapons delivery systems, facial recognition applications, policing, defense/strategic applications
Critical infrastructure	• Economy dependent on electricity grid • Human life requires functioning water and waste systems • High productivity and efficiency potential with AI systems • Complexity of systems and issue of progressively handing over control of critical infrastructure • Security of systems critical—AI introduces new threats/new solutions	Examples include optimizing operation and servicing of electricity, water, waste, internet/cell phone towers
Government	• AI systems can drive efficiency and productivity but generate multiple concerns • Ensuring systems don't threaten governing power (e.g., democratic values) • Concerns about "handing over control" of critical government systems (e.g., logistics, stockpiling, automating judicial decisions, electoral processes) • Bias in data resulting in discriminatory outcomes	Examples include decision-making and processing systems across logistics, employment, social security/welfare, payments, procurement, justice
Education/finance/healthcare	• AI outcomes in these sectors can have a material effect on people's quality of life • AI systems often have inherent data biases that render outcomes discriminatory/less effective • Extensive data privacy/data security concerns owing to sheer amount of personal data required to enable applications • Use of facial recognition—often required to enable applications but still necessary to define appropriate usage, e.g., regulation restricting continual surveillance in schools • Need for right to appeal certain decisions to a human	Examples include applications and delivery across all segments of education (from entry vetting systems to teaching and testing), finance (from loan approval to investment management) and healthcare (from diagnostic systems to surgical robots)

Source: StepStone Group analysis.

certain sectors. This is reflected in emerging regulations, which are already homing in on such "high-risk" sectors.

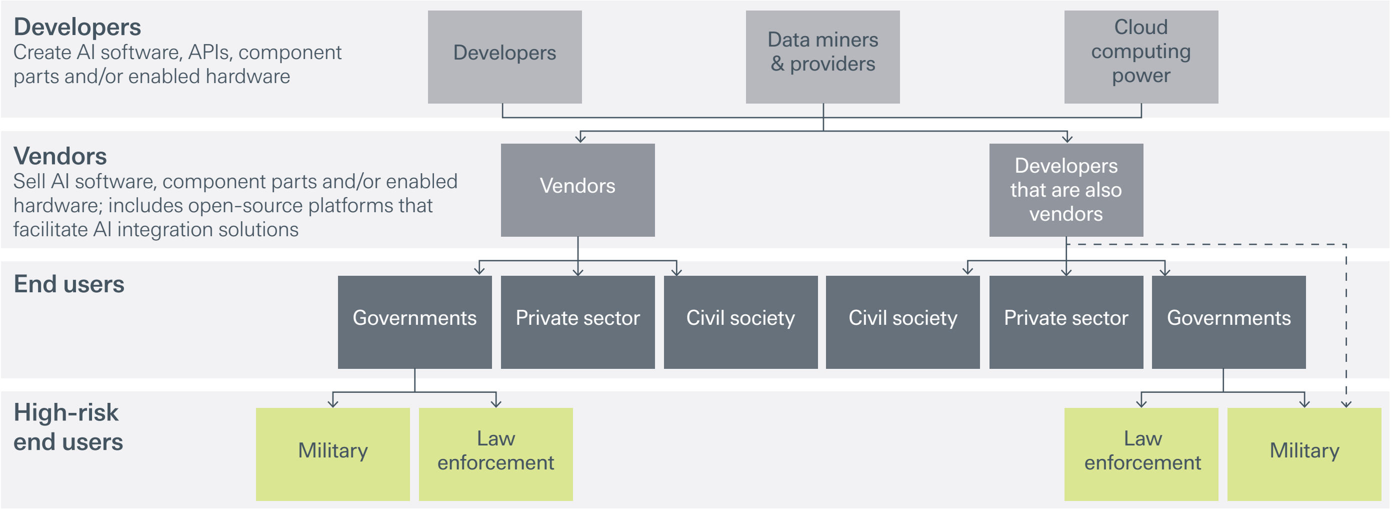
By recognizing that certain sectors pose greater AI-related risks, private market investors can begin preparing for the day when investments in these riskier sectors become more expensive to vet and monitor, are subject to greater regulatory scrutiny, and need to adequately compensate investors. Some types of assets may even need to be abandoned.

This sector lens is particularly important for GPs that invest in these sectors. To differentiate themselves, specialists will increasingly be required to have the skills and expertise to evaluate (pre-investment) and support (post-investment) AI development or integration in their portfolio companies.

- The healthcare, education and financial sectors receive significant funding from private equity, and the business models are already being significantly affected by newer AI-powered applications. One key aspect is that generative AI can facilitate personalization, which would in turn increase the value of offerings across these sectors—e.g., personalized drug formulation, individualized study systems and bespoke financial products.
- Similarly, AI in the real estate and infrastructure sectors is going to be an incredibly important consideration to drive increased build and operational efficiency. While the potential positive effects are huge, there are challenges related to the control and safety of such systems. Hacking an HVAC or water system can cause material harm; the same goes for manipulating government logistics or strategic stockpiles.
- Though relatively few GPs invest solely in the defense industry, today the relevant hardware or software systems are often being conducted within the ambit of other sectors and then applied in the defense sector or vice versa (a.k.a. "dual-use technology"). As such, GPs will need to understand the nature of *all* potential end users and whether the defense sector is likely to be a target customer. If it is, the company will need to meet a growing list of specific AI requirements from regulators and asset owners alike.

Another way to consider the risk profile is that the AI landscape contains multiple levels of users. As the risks and harms cascade down the layers, they affect these constituents differently.

FIGURE 8: LANDSCAPE OF AI DEVELOPERS, VENDORS AND USERS



Source: OECD, 2019.

The risks are generally highest for developers, and if those risks go unchecked, they percolate to end users where they can manifest as harms. Even if high-risk end users are aware of the risks, it is very difficult for them to manage said risks if responsible practices have not been embedded at the developer level and reinforced at the vendor level.

Ultimately, the agents in each layer carry a duty to drive Responsible AI practices. Only then will these responsible behaviors become self-reinforcing norms that propagate across the landscape.

Within the private markets, GPs tend to invest heavily in either developers or end-user applications or both. The ability to institute Responsible AI practices at the "top" of the cascade can go a long way in mitigating harm down the line. Owing to the nascency of many of the opportunities in generative AI, the venture capital community has a role to play in supporting entrepreneurs and founders by championing Responsible AI to alleviate upstream risk and minimize downstream harm. Consequently, the ability of GPs to offer value-add strategies to founders with respect to AI development best practices is becoming a consideration for LPs in manager selection.

Similarly, private equity or real asset fund managers are increasingly adopting applications enabled by generative AI. As such they need to ensure that third-party AI systems have been responsibly developed and are trustworthy. Vendor due diligence will be critical to such supply chain management. Failure to do so could result in significant commercial harm. Some assets will be stranded. As discussed earlier, supply chain management particularly in higher-risk sectors like critical infrastructure and government systems will be even more demanding. Again, LPs will be increasingly focused on how GPs are accounting for AI-related risks in their asset management practices.

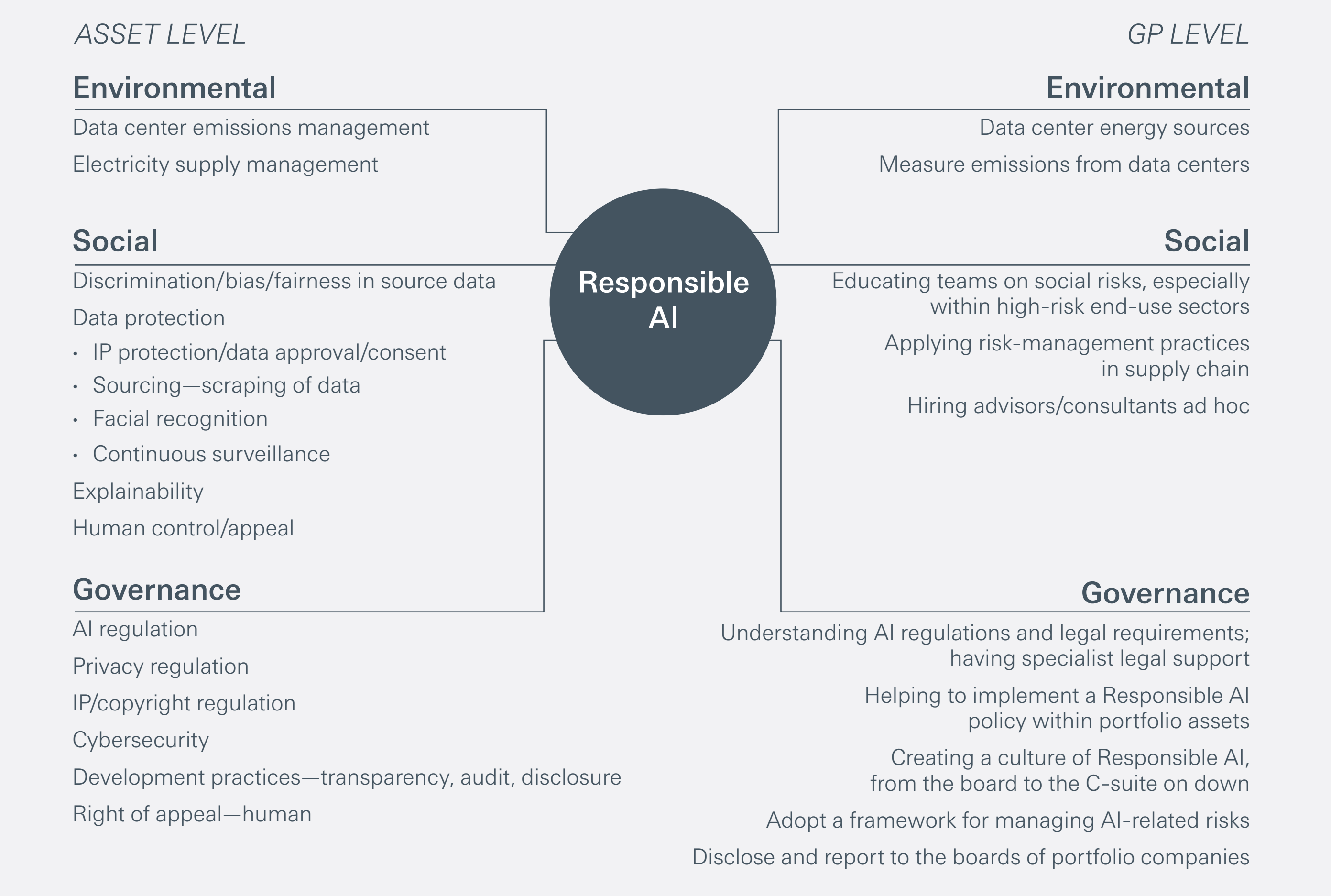
Approaches to addressing Responsible AI

Responsible AI integration at the Asset and GP level

Over the past decade, investors have become increasingly accustomed to integrating ESG considerations into their decision-making models to protect capital and add value. Though Responsible AI may appear to be another beast entirely, many of its concepts will nest neatly in GPs' and their underlying assets' existing ESG frameworks. This is shown in **Figure 9**, where key issues linked with Responsible AI have been parsed across E, S and G considerations. Having an ESG architecture in place will help GPs meet the challenges AI poses. Those that possess strong frameworks for dealing with social and governance issues may have a leg up, owing to the number of AI-related risks that reside within those buckets. Still, that may not be enough.

- **Pre-investment**, GPs will focus on determining how developed the asset is with respect to Responsible AI practices. Evidence of unmitigated risks—particularly across social considerations and an immature governance function—could create an opportunity for GPs to add value. In some cases, they may have to bring on consultants; in others, they may need to walk away. Identifying third-party partners might be challenging: Audit firms, cybersecurity specialists and software engineers won't be able to evaluate the entire AI system. Most of the applications will be new, and no one has all the skills or bona fides. But like anything

FIGURE 9: NESTING RESPONSIBLE AI IN ESG FRAMEWORKS



ASSET LEVEL

Environmental

Data center emissions management

Electricity supply management

Social

Discrimination/bias/fairness in source data

Data protection

- IP protection/data approval/consent
- Sourcing—scraping of data
- Facial recognition
- Continuous surveillance

Explainability

Human control/appeal

Governance

AI regulation

Privacy regulation

IP/copyright regulation

Cybersecurity

Development practices—transparency, audit, disclosure

Right of appeal—human

Responsible AI

GP LEVEL

Environmental

Data center energy sources

Measure emissions from data centers

Social

Educating teams on social risks, especially within high-risk end-use sectors

Applying risk-management practices in supply chain

Hiring advisors/consultants ad hoc

Governance

Understanding AI regulations and legal requirements; having specialist legal support

Helping to implement a Responsible AI policy within portfolio assets

Creating a culture of Responsible AI, from the board to the C-suite on down

Adopt a framework for managing AI-related risks

Disclose and report to the boards of portfolio companies

Source: StepStone Group analysis.

else, we'd expect a cottage industry of system specialists to develop over time.

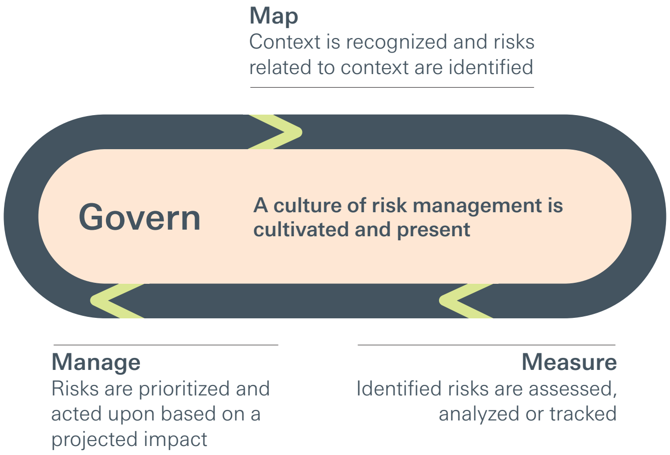
- **Post-investment**, GPs will find opportunities to add value by instilling portfolio assets with improved risk-management and governance frameworks. Accretion through good governance is often considered a strength of private market GPs. Though the details may differ, Responsible AI should not be any different.

To help asset owners and GPs weave Responsible AI into their ESG frameworks, in the next section, we offer some practical advice drawing from NIST's AI Risk Management Framework. As with any risk-management framework, governance is foundational.[5] It starts with the board of directors and trickles down through the ranks.

Laying down a culture of responsible behaviors is as important for start-ups as it is for large corporations. For different reasons, it is challenging for both types of organizations. In start-ups, someone may be a founder, director, CEO and software engineer all rolled into one. The task of building a responsible culture can easily find itself on the proverbial backburner. For large organizations, the challenge is one of communication, consistency and preventing silos. Critically, implementing Responsible AI practices is not the responsibility of "an ESG person" or the compliance department. It is a shared responsibility, which flows from the board.

For GPs this means developing the appropriate governance structures across their assets to ensure that Responsible AI is on the board's agenda and executive management is willing and able to implement it. Again, this is not an ESG issue, nor does it fall solely to a GP's ODD or compliance team. Understanding how GPs implement Responsible AI throughout their portfolios will be paramount for LPs.

FIGURE 10: NIST AI RISK-MANAGEMENT FRAMEWORK


Map
Context is recognized and risks related to context are identified
Govern
A culture of risk management is cultivated and present
Manage
Risks are prioritized and acted upon based on a projected impact
Measure
Identified risks are assessed, analyzed or tracked

Source: NIST, 2023.

Drawing again from NIST's framework, we've summarized the points we find most relevant for GPs to ensure that company management is addressing Responsible AI in a systematic manner. Furthermore, GPs will need to consider how they monitor risks and benefits at an asset and portfolio level (**Figure 11**).

[5] Throughout this section, governance refers to the function within an organization.

FIGURE 11: HIGHLIGHTS OF NIST'S FRAMEWORK FOR GPS AND LPS

Governance	
Align with legal and regulatory requirements	Institute process on decommissioning of AI systems
Integrate Responsible AI practices into policy suite	Ensure training is delivered to enable culture of risk awareness, accountability for AI systems
Determine who is accountable for AI systems	Document risks and mitigation strategies
Determine how AI risk management will be executed (accountability, capacity, capability)	Managing AI risks associated with third-party systems including failures of such systems
Conduct inventory of AI systems	Ensure transparency and disclosure through organization on AI risk management

Map	
Context and purpose of the AI system is understood	Mapping benefits/positive impact driven by system and understanding nature and limits of AI system
Mapping of risks within overall AI system including third-party AI	Compliance with relevant technical standards/certification
Consider how new risks will be identified over time	Establish feedback process from actors in AI system
AI system requirements to manage identified risks, e.g., what needs to be done to ensure data rights are respected if data rights pose an identified risk	

Measure	
Determine how to measure that the system is meeting the purpose it was designed for and is delivering valid and reliable output	Implement best-practice test, evaluation, verification and validation (TEVV) procedures on model
Determine how to measure AI risks (identified in mapping process). If risks can't be measured, then how to monitor them?	Set up independent assessors/reviewers (internal/external) of model
Appropriate documentation of model, and of risks	Determine if there is sufficient feedback coming into system

Manage	
Prioritization and response to risks identified	Procedures to manage previously unknown risks
Relevant resources available to manage risks	Monitoring of third-party systems—are they delivering or producing unintended/excessive risks?
Manage responding to incidents and errors created by system	Actively shut off systems as required

Sources: NIST, 2023; StepStone Group analysis.

- **Mapping** relates to understanding the system's purpose and diagramming the risks, impacts, costs and benefits. This applies to internally built as well as third-party systems.
- **Measuring,** in contrast, is focused on determining how to actively monitor the risks and benefits.
- **Management** uses that data to realize AI's benefits and to mitigate its risks.

Just like their assets, GPs will need to invest in human capital, hiring specialists or training their investment professionals as the situation dictates to ensure Responsible AI is part of the investment due diligence process, investment-period value creation and post-investment monitoring.

There will, no doubt, be many opportunities for GPs to apply generative AI applications within their own practices: industry analysis; data compilation; target outreach processes; and investment memos, among others, are all likely to be enhanced. Managing data privacy and security will need to be enhanced particularly considering the sensitivity of data being shared with open generative AI models.

Responsible AI integration at the LP level

Ultimately an LP is concerned that a GP appropriately identifies both risks and benefits (discussed extensively in prior sections) pre-investment and has plans to address both post-investment. This comes back to the principles of value protection and enhancement that are foundational to responsible investment practices.

Pre-investment, Responsible AI considerations will need to be embedded in the LP DDQ to get a measure of GP awareness and capability in this regard. As noted, this should be adapted depending on whether the GP is driving the owner of a developer versus a user, and the extent of the GP's likely exposure to high-risk sectors. Evaluation needs to be further calibrated to how material Responsible AI considerations will be to the GP's specific strategy and also recognition of where the GP is in terms of maturity of this topic.

To this end an LP may consider a GP capability suite—both internal and external. Does the GP have a relevant suite of external specialist consultants assembled? Does the GP have awareness of and capability/support to navigate the regulatory landscape? Is the GP developing playbooks to support founders on implementation? How is the GP conducting vendor due diligence pre-investment? How will this be adjusted post-investment? How is the GP going to measure the effectiveness of their Responsible AI practices across their portfolio?

At StepStone, we are moving to include such considerations in our due diligence processes. We are particularly focused on the exposure to such risks through primary and co-investment strategies. We strongly believe that considerations around Responsible AI are investment concerns—seeking value accretion and protection for our clients. We believe that having a strong responsible investment foundation helps us better evaluate these opportunities.

Dominant AI principle frameworks

AI principle frameworks are the first salvo in trying to ensure that AI systems are trustworthy. UNESCO and the OECD have been particularly active in framework development. Many countries have developed similar frameworks that can trace their roots back to the work of these two intergovernmental organizations.

Overall, there is a convergence around what is desirable—with the starting point being a consensus on the need for "no harm to humans." See **Figure 12**. Where divergence arises is on the interpretation of certain values-based topics that, as discussed on **page 7**, fall within the purview of Ethical AI: for example, individuals' control over their data, controls over continual surveillance.

Similarly, while there is broad desire for global cooperation, it should be noted that the developed world dominates the "negotiating table." The complex relationship between China and OECD members also features heavily. No country wants to be left behind in the AI arms race; global cooperation is seen as key to supercharging the AI ecosystem. However, the divergence on certain ethical considerations may prevent this effort from being truly global.

FIGURE 12: SUMMARY OF GLOBAL AI FRAMEWORKS

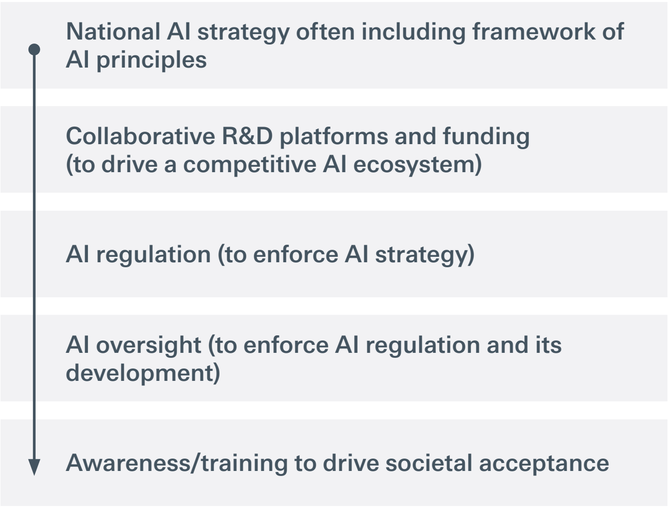
	US— AI bill of rights	OECD	UNESCO	China including Global AI Governance Initiative
Effective outcomes/aligned with objectives	X	X	X	
Independent evaluation of system				
Disclosure of harm	X	X	X	X
No discrimination in data or models	X	X	X	X
No abusive data practices—collection/ invasive practices/clear consent	X	X	X	X
Individuals have agency over own data	X	X	X	
Disclosure of data/data decisions/ auditable/traceable	X		X	X
Notice and explanation—provide useful explanation of outcome/transparency/ explainability	X	X	X	X
Human alternatives/oversight—appeal to human for high-risk/adverse decisions	X		X	X
Safe to community/do no harm with output/ social security	X	X	X	X
No continuous surveillance at home, education, work	X			X (Clear allowances for government)
Inclusive growth/sustainable development		X	X	X
Human-centered values/social sustainability/ fairness		X	X	
Awareness and literacy for users and developers/upskilling			X	X
International cooperation		X	X	X

Source: StepStone Group analysis.

Global regulation

Building on the foundational AI principle frameworks, governments have begun to construct the AI policy landscape. As of December 2023, there have been over 1,000 policy initiatives from 69 countries. Most have focused on developing national AI strategies, regulatory policies or oversight bodies. Beyond these, public consultations, awareness campaigns and collaborative R&D platforms have received the most funding and attention.

FIGURE 13: TYPICAL POLICY PROGRESSION



National AI strategy often including framework of AI principles

Collaborative R&D platforms and funding (to drive a competitive AI ecosystem)

AI regulation (to enforce AI strategy)

AI oversight (to enforce AI regulation and its development)

Awareness/training to drive societal acceptance

Source: StepStone Group analysis.

Most agree that China and the European Union are the leaders in AI policy. The EU focuses on making AI "ethical and secure by design" and prioritizes international and interdisciplinary collaboration, which it hopes to achieve by building world-class data centers and AI hubs. Experimentation, innovation and broad government support/adoption figure heavily into the EU's plans.

China's policy priorities have quickly evolved. Its early-move policy position announced in 2017 set a goal of global AI leadership. This was amended in 2019 as concerns arose about how the Chinese government could exert some control over private-sector AI development and the emergent AI abuses, which were creating social discontent. China then embraced AI trustworthiness as its core policy objective. More recently, in 2023, China took measures that align it with the EU, the OECD and the US.[6]

By contrast, the US's development of an overarching policy began relatively late. While the AI Bill of Rights (2023) is less comprehensive than other national frameworks, some of the US's subsequent efforts are considered to be world leading. NIST's AI Risk Management Framework is one such example.[7]

[6] See, for example, the Global AI Governance Initiative.
[7] Refer to page 14 for more detail.

China's AI lead

China has emerged as the early leader in the AI arms race. It was the first country to create a national AI strategy, and no country has spent more on AI.

Its first policy, which came in 2017, sought to quickly bring China up the learning curve and to make it the world leader in AI by 2030. The Chinese government reckoned its domestic AI industry would one day be worth RMB 1 trillion (€130 billion). To get there, the government would partner with domestic tech companies and build a state-of-the-art technology park for AI research.

In 2019, concerns began to foment. Domestically, the government grew worried about how the pace of development might imperil national security. The international community was concerned the Chinese government might abuse AI. So China laid out ethical norms and began working on building trustworthy AI systems.

Since then, China has become the apparent champion of global collaboration. It is aligning itself with the OECD's AI framework and has launched the Global AI Governance Initiative, which calls for "equal rights" when developing AI regardless of how large or influential a country might be, to take one such example.[8] Still, many are skeptical.

There is inherent tension between governance and national security. One case study is the continual use of facial recognition technology (FRT) in public spaces. FRT captures unique biometric information, which is the most personal of personal data and as such should be covered under data protection laws.

For the past two decades, China has been at the forefront of both the development and deployment of FRT. The country is also a major exporter of the technology, which means that FRT advances in China have global impact. As of 2023, there were reportedly over 700 million cameras deployed.[9] Use spans Tianjin Railway implementing FRT in its payment systems, enabling people to ride the metro without having to purchase tickets beforehand; universities use FRT to track attendance, reduce absenteeism, identify inattentive students and verify the identity of people taking college entrance exams. FRT has also been used to prevent children from playing computer games between 10 p.m. and 8 a.m., which is prohibited by the Chinese government.

In 2021 China explicitly classified biometric information—including facial information—as personal information. It also enacted the Personal Information Protection Law (PIPL) and the Data Security Law.[10]

The issue is that all these procedures can be sidestepped in the name of public security, which can be interpreted quite expansively. There is some evidence that FRT has been deployed to police individuals, including for minor infractions such as jaywalking. The penalties for abuse are also asymmetric—being relatively light for government and heavier for private companies.[11]

Navigating a course of data protection and national security concerns is being played out in countries around the world. Observing China's policy development path can be instructive for other nations.

[8] Global AI Governance Initiative (mfa.gov.cn)

[9] Dashveenjit Kaur. 2023. "After years of dominating facial recognition technology, China is ready to govern it." *Tech Wire Asia*, August.

[10] The PIPL specifically classifies biometric information as a primary type of "sensitive personal information," and emphasizes that such sensitive personal information can only be processed with the individual's consent, for a specific purpose, with sufficient necessity requiring a prior risk assessment. The Data Security Law emphasizes that data-related activities must be "conducive to economic and social development, promote people's well-being, and comply with social morality and ethics."

[11] Article 68 of the PIPL indicates that violation of personal data rights by the government only leads to administrative liabilities, which would rely on self-correction measures conducted by state agencies.

FIGURE 14: EXAMPLES OF EXISTING AND EMERGING AI-SPECIFIC REGULATORY APPROACHES

Jurisdiction	Legislation and regulation	Standards	Principles
Canada	• Directive on automated decision-making (2019) • Proposed Bill C-27, Digital Charter Implementation Act, including AI and Data Act (AIDA) (2022)	Proposed CAN-ASC-6.2: Accessible and Equitable AI Systems (2023)	Canada's Digital Charter (2019)
United Kingdom	• Proposed Online Safety Bill (2022) • Proposed Data Protection and Digital Information Bill (2023)	Algorithmic Transparency Standard (Central Digital Data Office, 2021)	A pro-innovation approach to AI regulation (2023)
United States	• Federal Trade Commission Act, for deceptive practices from deepfakes or chatbots (1914) • Proposed Algorithmic Accountability Act (US AAA) (2022)	NIST AI Risk Management Framework (2023)	Blueprint for an AI Bill of Rights (2023)
European Union	• Proposed EU AI Act (2021) • Proposed updates to the EU Product Liability Directive (2022) • Proposed AI Liability Directive (2022) • EU's Digital Services Act (2022)	CEN/CENELEC standards for AI and related data (forthcoming)	Ethics guidelines on AI (2018)
Brazil	• Report and proposed substitute text for the draft bills 5051/2019, 21/2020 and 872/2021 (2022) • Proposed Bill 705 on the compatibility of AI use in the public sector with ESG practices (2022)	Incorporation of international standards and national standards by the Brazilian Association of Technical Standards (ABNT)	Proposed Art. 3 of the proposed substitute text for draft bills 5051/2019, 21/2020 and 872/2021 (2022)
China	• Chinese Internet Information Service Algorithmic Recommendation Provisions (2021) • Opinion on Strengthening the Ethics and Governance of Science and Technology (2022)	National Standards for Autonomous Vehicle Testing (2018)	• New Generation AI Ethics Specifications (2019) • New Generation AI Code of Ethics (2021) • White Paper on Trustworthy AI (2021) • Internet Information Service Algorithmic Recommendation Management Provisions (2021)
Intergovernmental organizations	Proposed Council of Europe Convention on AI, Human Rights, Democracy and the Rule of Law (2023)	• ISO 31000 Risk management (2009, 2018) • ISO/IEC 23053:2022 Framework for AI Systems Using Machine Learning (ML) (2022)	• OECD Recommendation of the Council on AI (2019) • UNESCO Recommendation on the Ethics of AI (2021)

Source: StepStone Group analysis.

Conclusion

Over the past two decades, we have seen both harms and benefits emerge from the internet's widespread adoption. Drawing on this knowledge, as society is embracing the incredible opportunity that the AI complex presents, it is moving swiftly to address emerging harms, especially within generative AI models. The emergence of AI principles around the world reflects societal ambitions and is an important precursor to the policy and regulatory landscape that is being assembled. All this is placed against the backdrop of how strategically important this technology will be in shaping economies. AI is the new arms race, and there is tension around how best to govern it responsibly while allowing it to flourish and drive growth. Asset owners and GPs have an opportunity to drive Responsible AI practices through their assets. This will be critical to ensuring the incredible value opportunity is not scuttled. Private markets appear once again suited to drive enhanced governance and operational practices—both of which align with GPs' core skill sets. But this is not a cost-free exercise: capacity will need to be built; systems and processes, enhanced. GPs need to mitigate the risks because assets that get mired in the myriad of AI risks will at best result in extended hold periods and high cost bases. On the flip side the value accretion opportunities are legion. Some GPs and asset owners will focus on the societal impacts that can be driven through Responsible AI practices, arguing for real-world outcomes and change. This is an important lens but it is not required for mainstream adoption of Responsible AI. Everyone has much to gain from this incredible opportunity. Responsible AI can help us make the most of it.

Reference links

US Artificial Intelligence Safety Institute

MLCommons AI Safety

AI Alliance

Partnership on AI

The Santa Clara Principles on Transparency and Accountability in Content Moderation

UNESCO Global AI Ethics and Governance Observatory

Blueprint for an AI bill of rights

AI Standards Hub

The Global Partnership on Artificial Intelligence

OECD AI Catalogue of Tools & Metrics for Trustworthy AI

The Language of Trustworthy AI: An In-Depth Glossary of Terms

Map of Global AI Regulations

This document is for informational purposes and is meant only to provide a broad overview for discussion purposes. This document does not constitute an offer to sell, a solicitation to buy, or a recommendation for any security, or as an offer to provide advisory or other services by StepStone Group LP, StepStone Group Real Assets LP, StepStone Group Real Estate LP, StepStone Group Private Wealth LLC, StepStone Group Private Debt AG, StepStone Group Europe Alternative Investments Limited and StepStone Group Private Debt LLC, their subsidiaries or affiliates (collectively, “StepStone”) in any jurisdiction in which such offer, solicitation, purchase or sale would be unlawful under the securities laws of such jurisdiction. The presentation is being made based on the understanding that each recipient has sufficient knowledge and experience to evaluate the merits and risks of investing in private market products. Information contained in this document should not be construed as financial or investment advice on any subject matter. StepStone expressly disclaims all liability in respect to actions taken based on any or all of the information in this document. This document is confidential and solely for the use of StepStone and the existing and potential investors or clients of StepStone to whom it has been delivered, where permitted. By accepting delivery of this presentation, each recipient undertakes not to reproduce or distribute this presentation in whole or in part, nor to disclose any of its contents (except to its professional advisors), without the prior written consent of StepStone.

Expressions of opinion are intended solely as general market commentary and do not constitute investment advice or a guarantee of returns. All expressions of opinion are as of the date of this document, are subject to change without notice and may differ from views held by other businesses of StepStone.

Some information used in the presentation has been obtained from third parties through various published and unpublished sources considered to be reliable. StepStone does not guarantee its accuracy or completeness and accepts no liability for any direct or consequential losses arising from its use. Thus, all such information is subject to independent verification by prospective investors.

All information provided herein is subject to change.

All valuations are based on current values calculated in accordance with StepStone’s Valuation Policies and may include both realized and unrealized investments. Due to the inherent uncertainty of valuation, the stated value may differ materially from the value that would have been used had a ready market existed for the portfolio investments or a different methodology had been used. The long-term value of these investments may be lesser or greater than the valuations provided.

StepStone Group LP, its affiliates and employees are not in the business of providing tax, legal or accounting advice. Any tax-related statements contained in these materials are provided for illustration purposes only and cannot be relied upon for the purpose of avoiding tax penalties. Any taxpayer should seek advice based on the taxpayer’s particular circumstances from an independent tax advisor.

Prospective investors should inform themselves and take appropriate advice as to any applicable legal requirements and any applicable taxation and exchange control regulations in the countries of their citizenship, residence or domicile which might be relevant to the subscription, purchase, holding, exchange, redemption or disposal of any investments. Each prospective investor is urged to discuss any prospective investment with its legal, tax and regulatory advisors in order to make an independent determination of the suitability and consequences of such an investment.

An investment involves a number of risks and there are conflicts of interest. Please refer to the risks and conflicts disclosed herein or in relevant disclosure documents associated with potential investments.

Each of StepStone Group LP, StepStone Group Real Assets LP, StepStone Group Real Estate LP, StepStone Group Private Wealth LLC and StepStone Group Private Debt LLC is an investment adviser registered with the Securities and Exchange Commission (“SEC”). StepStone Group Europe LLP is authorized and regulated by the Financial Conduct Authority, firm reference number 551580. StepStone Group Europe Alternative Investments Limited (“SGEAIL”) is an investment adviser registered with the SEC and an Alternative Investment Fund Manager authorized by the Central Bank of Ireland and StepStone Group Private Debt AG (“SPD”) is an SEC Exempt Reporting Adviser and is licensed in Switzerland as an Asset Manager for Collective Investment Schemes by the Swiss Financial Markets Authority FINMA. Such registrations do not imply a certain level of skill or training and no inference to the contrary should be made.

In relation to Switzerland only, this document may qualify as “advertising” in terms of Art. 68 of the Swiss Financial Services Act (FinSA). To the extent that financial instruments mentioned herein are offered to investors by SPD, the prospectus/offering document and key information document (if applicable) of such financial instrument(s) can be obtained free of charge from SPD or from the GP or investment manager of the relevant collective investment scheme(s). Further information about SPD is available in the SPD Information Booklet which is available from SPD free of charge.

All data is as of July 2024.

We are global private markets specialists delivering tailored investment solutions, advisory services, and impactful, data-driven insights to the world's investors. Leveraging the power of our platform and our peerless intelligence across sectors, strategies, and geographies, we help identify the advantages and the answers our clients need to succeed.

For more information regarding StepStone's research, please contact us at research@stepstonegroup.com.

STEPSTONE
stepstonegroup.com