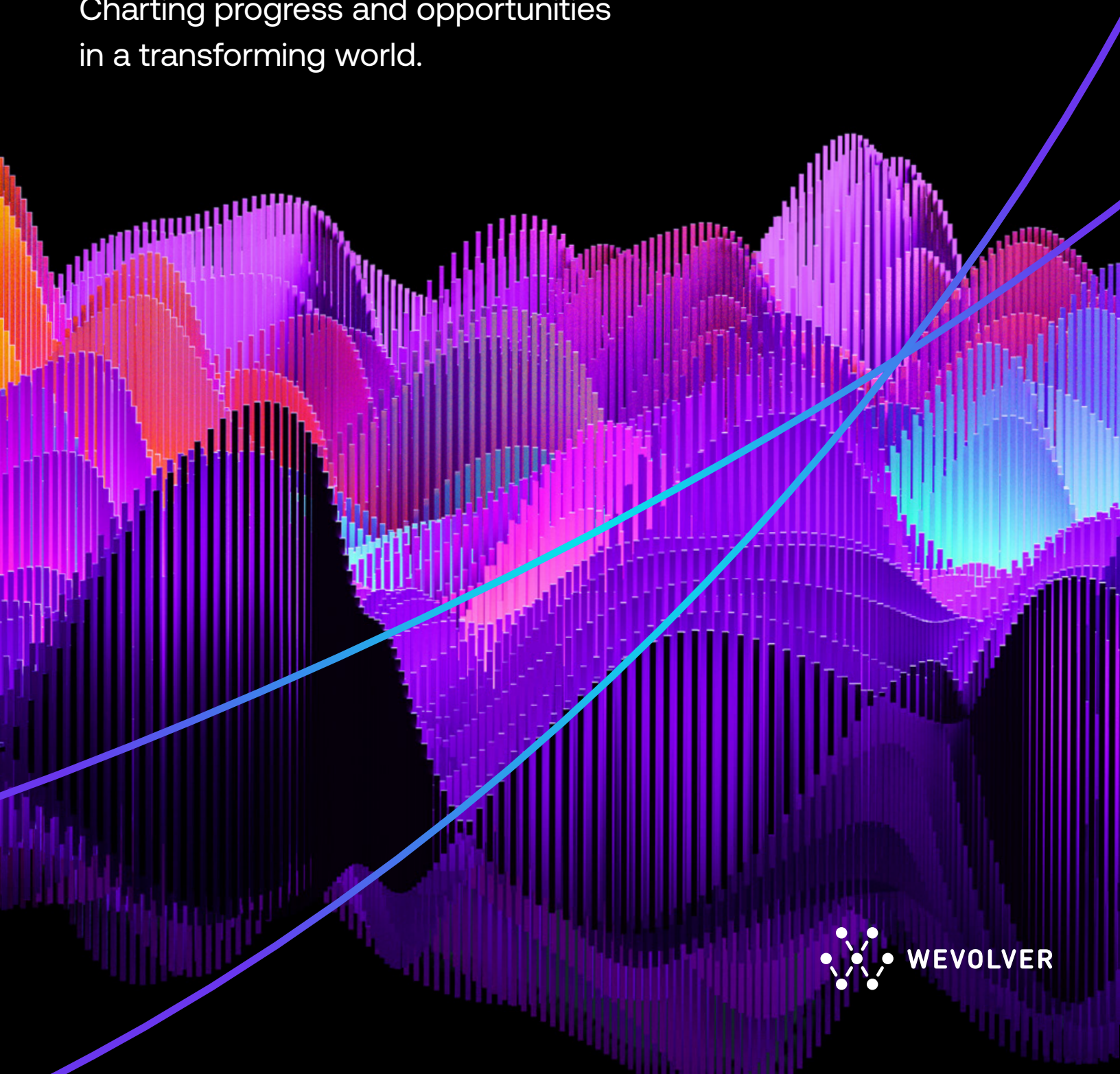
arm

Arm AI Readiness Index

Charting progress and opportunities
in a transforming world.



WEVOLVER

CONTENT

Foreword

WILL ABBEY,
EXECUTIVE VICE PRESIDENT AND CHIEF COMMERCIAL OFFICER, ARM

Throughout history, certain technological advancements have fundamentally altered our world. The printing press democratized knowledge. The industrial revolution transformed manufacturing. The internet connected humanity in ways previously unimaginable. Today, we stand at another such pivotal moment with artificial intelligence.

Like those historic inflection points, artificial intelligence (AI) is not merely an incremental improvement on what came before; it represents a paradigm shift that is redefining how we work, communicate, create, and solve problems. And like those earlier transformations, the full impact of AI will be determined not just by the technology itself, but by how we choose to develop, deploy, and govern it.

What was once considered the realm of science fiction has become an integral part of our daily lives and business operations. The AI Readiness Index report represents a comprehensive effort to understand where we stand today in a rapidly transforming world and chart a course for the future.

As we navigate this transformative era, one thing has become abundantly clear: AI is no longer a technology of tomorrow—it is the driving force of now. Our research reveals that 82% of global business leaders are already using AI applications, with overwhelming majorities expecting to increase their investments in the coming years. Yet despite this widespread adoption, only 39% report having a clearly defined, comprehensive strategy in place.

This gap between adoption and strategic implementation underscores the complex challenges organizations face. From building robust technical

foundations and addressing security concerns to navigating evolving regulatory landscapes and developing necessary talent, the path to AI success requires careful consideration of multiple interconnected factors.

What makes this moment particularly significant is the shifting nature of the conversation itself. We have moved beyond asking whether AI will transform our industries to asking how we can responsibly harness its potential to drive innovation, efficiency, and growth. The focus has rightfully expanded from purely technical considerations to encompass broader implications for security, sustainability, governance, and workforce development.

The AI Readiness Index aims to serve as both a mirror and a compass—reflecting the current state of AI readiness across industries and regions while providing direction for organizations at various stages of their AI journeys. By combining extensive survey data from 665 business leaders with expert analysis and case studies, we offer a nuanced view of the challenges and opportunities that lie ahead.

As you explore the findings in this report, I encourage you to consider not just where your organization stands today, but where it needs to be tomorrow. The organizations that will thrive in this new era will be those that approach AI with both ambition and responsibility—embracing its transformative potential while thoughtfully addressing the technical, ethical, and human dimensions of implementation.

The future of AI is being written today, by leaders willing to ask difficult questions, take bold risks and make strategic investments in the foundations of tomorrow's success. My hope is that this report will serve as a valuable resource as you write your organization's chapter in that future.

CHAPTER 1

The Global State of AI Readiness: A Data-Driven Analysis

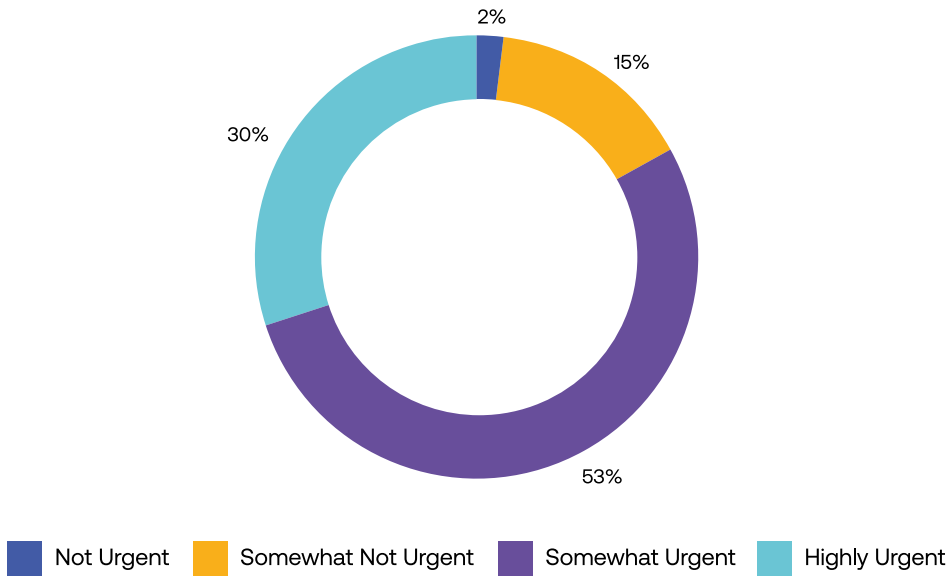
METHOD
COMMUNICATIONS

In boardrooms and strategy sessions across the globe, artificial intelligence has moved from a distant aspiration to an immediate priority. Our extensive survey of business leaders reveals a notable reality: AI adoption is accelerating rapidly, with 82% of global businesses already deploying AI applications in their day-to-day operations.. This broad adoption of artificial intelligence spans industries and regions, demonstrating that AI is no longer the exclusive domain of tech giants but has become an essential technology for enterprises of all types.

Customer service stands at the forefront of this adoption wave, with 63% of organizations deploying AI to enhance customer interactions. Document processing (54%), IT operations (51%), and security applications (51%) follow closely behind, showing how AI has rapidly infiltrated core business functions. This isn't merely experimentation—it's transformation at scale.

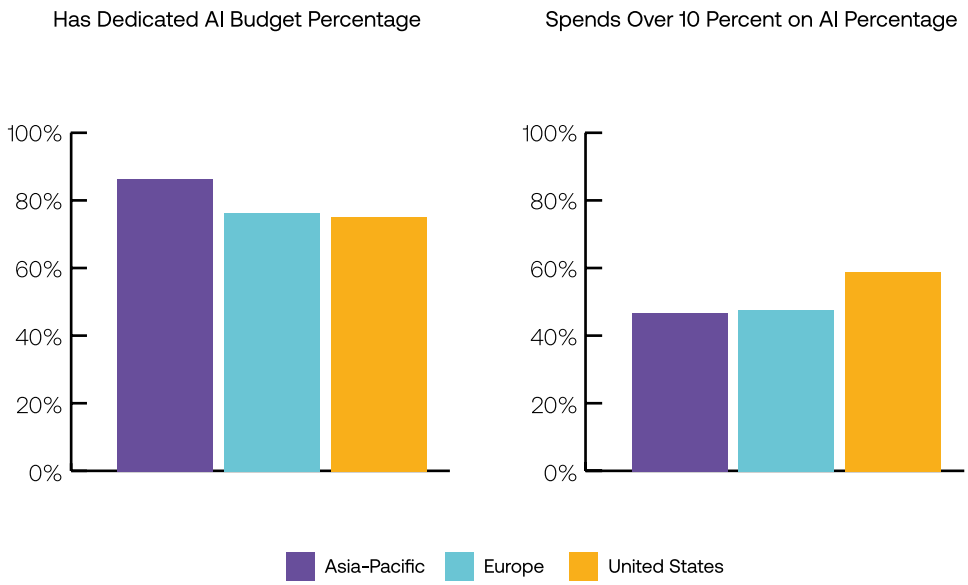
The enthusiasm for AI runs deep within organizational hierarchies. An overwhelming 90% of leaders report their organizations are receptive to AI-driven changes, with 83% considering AI adoption "urgent."

83% of Global Leaders Believe it is Urgent for their Organizations to Embrace AI; with 30% Saying it's Highly Urgent



Most tellingly, 82% of businesses have secured executive sponsorship at the highest levels, with CEOs and senior management personally championing AI initiatives. This top-down commitment signals that AI has transcended technical curiosity to become a strategic imperative.

Financial commitments reinforce this strategic shift. Eight in ten organizations have established dedicated AI budgets, with regional variations highlighting different approaches to investment. The Asia-Pacific region leads in budget allocation, with 86% of APAC businesses dedicating resources to AI initiatives, compared to 76% in Europe and 75% in the United States. However, American organizations are investing more aggressively, with 57% committing more than 10% of their IT budgets to AI, surpassing European (46%) and APAC (45%) counterparts.



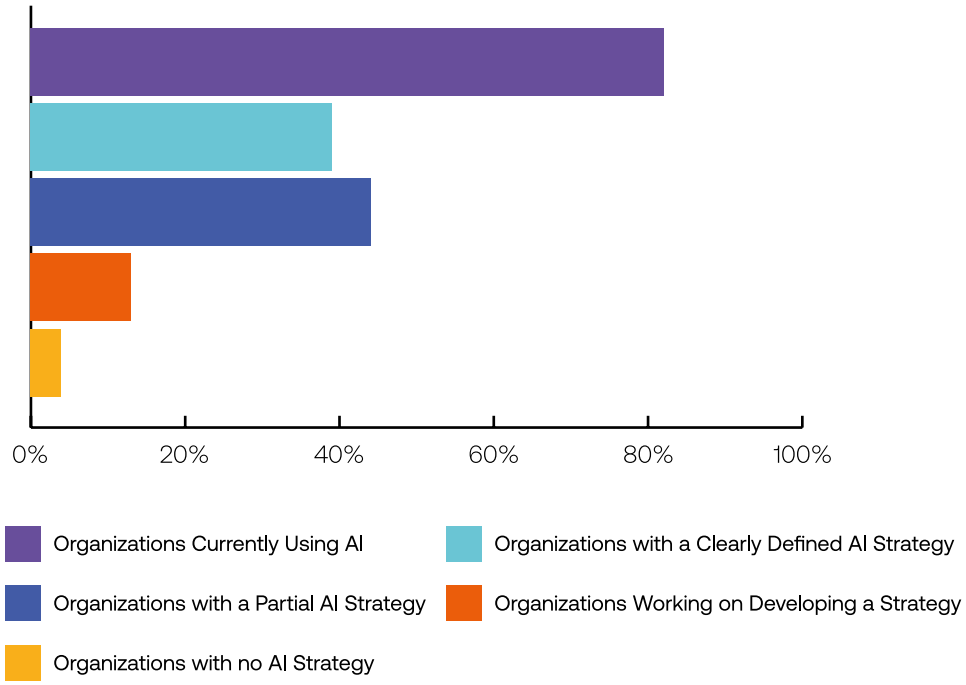
This financial commitment shows no signs of wavering, as 87% of business leaders anticipate increasing their AI budgets over the next three years.

"AI will eliminate repetitive work, allowing employees to engage in more strategic and creative tasks." - *Business leader survey respondent*

What's driving this rush toward AI? In a word, efficiency. A compelling 63% of global leaders identify operational efficiency as the primary benefit they expect from AI, with 80% making it the central focus of their 2025 AI strategy. As one business leader articulated, "AI will eliminate repetitive work, allowing employees to engage in more strategic and creative tasks." Another explained how "AI advancements will impact our organization by enhancing efficiency in decision-making and innovation across operations while requiring continuous adaptation to new technologies and regulatory changes."

Yet beneath this momentum lies a concerning paradox.

Adoption-Strategy Paradox



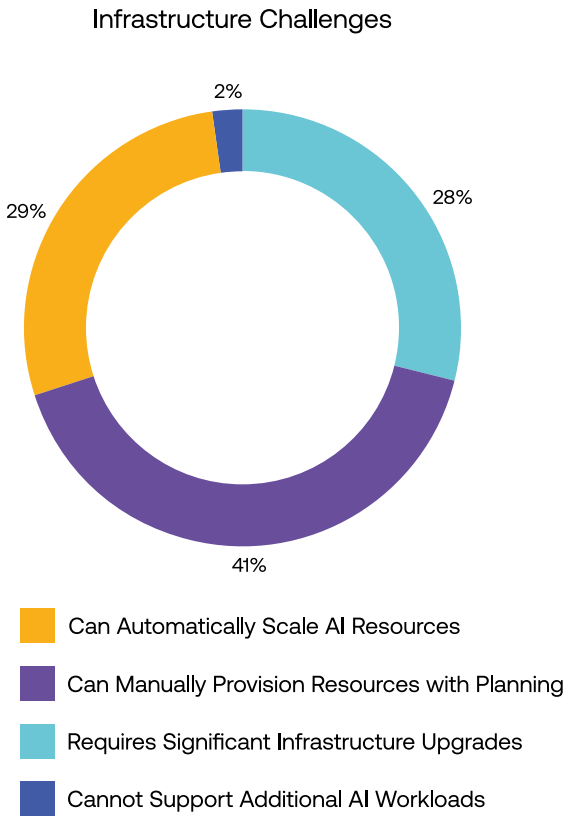
Despite widespread adoption and executive support, only 39% of organizations have developed a clearly defined, comprehensive AI strategy. Even fewer—just 37%—have implemented a robust change management plan to guide AI implementation across their businesses. This gap underscores that while AI is deployed, many organizations may not fully understand how to integrate it effectively or ready their workforce for the changes ahead.

One Vice President in the financial industry's IT department acknowledged: "We have our work cut out for us in formalizing these toolsets. Their use has grown organically within the business units, so we have duplication. I think we're getting there, and that's certainly part of my role—to figure that out."

This lack of a cohesive strategy raises critical questions about the sustainability of current AI initiatives and their long-term impact on business performance.

PROCEED WITH CAUTION

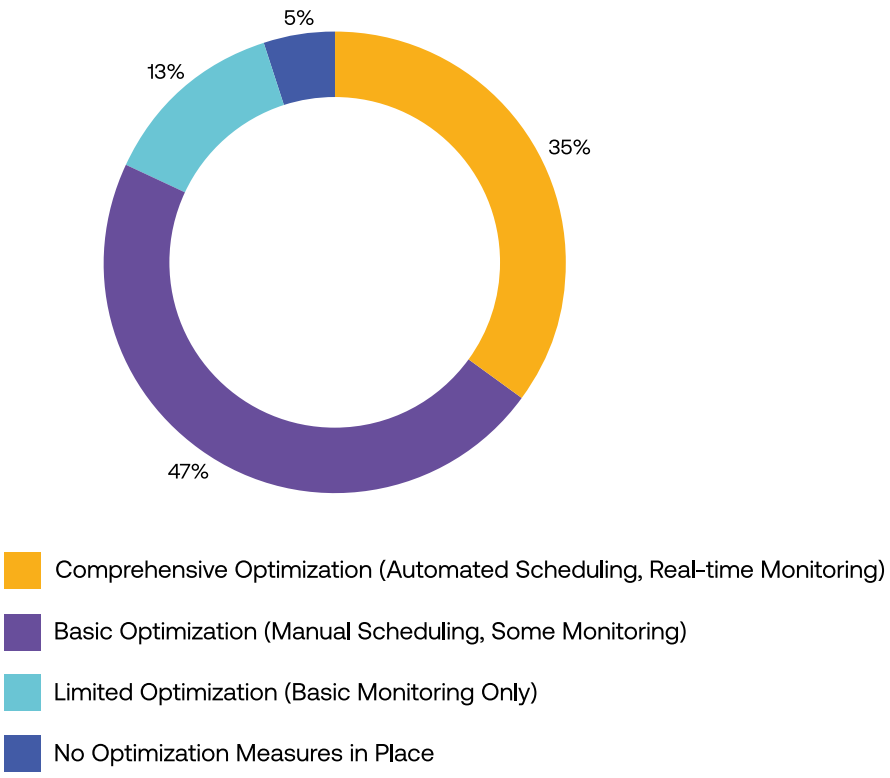
The enthusiasm for AI must be tempered with a sobering reality: many organizations remain ill-equipped to scale their AI initiatives effectively. Our research reveals significant shortcomings in three critical areas—**infrastructure readiness, talent availability**, and **data quality**—each representing a potential barrier to realizing AI's full potential.



Infrastructure limitations are particularly concerning. Only 29% of organizations have systems or storage resources to meet growing AI

demands. Even fewer—a mere 23%—have dedicated power infrastructure to handle the increased energy requirements of AI workloads. This means **77% of businesses have only begun upgrading their facilities**—if they have any power management strategy. Additionally, 65% of leaders acknowledge that their organizations lack comprehensive energy efficiency optimization measures for AI systems, raising questions about operational costs and environmental impact as these workloads grow.

AI Efficiency Optimization Level



The talent gap poses an equally significant challenge. A third of business leaders (34%) report their organizations are “significantly under-resourced” or “under-resourced” when it comes to AI expertise, while nearly half (49%) identify a lack of skilled talent as a primary barrier to successful AI implementation. As one C-level executive in manufacturing lamented, “It’s about manpower talent. It’s so hard to find people who know AI and can talk about AI.”

Despite recognizing the talent gap, organizations are taking inconsistent approaches to addressing it. While two-thirds (66%) of leaders express intentions to upskill existing employees to adapt to increased AI integration, 39% still don't have dedicated programs to develop AI skills among their workforce. This disconnect between recognized need and decisive action threatens to widen the talent gap rather than close it.

Perhaps most fundamentally, organizations struggle with data readiness—the essential foundation for effective AI applications. Nearly half of businesses leverage customer data (49%) and operational data (48%) for AI initiatives. However, data management practices remain largely rudimentary. Only about half (53%) of organizations have basic data automation processes for AI/ML models, while 18% rely on manual or ad hoc data cleaning procedures. Not surprisingly, 46% of leaders cite data quality and accessibility issues as a major barrier to successful AI implementation.

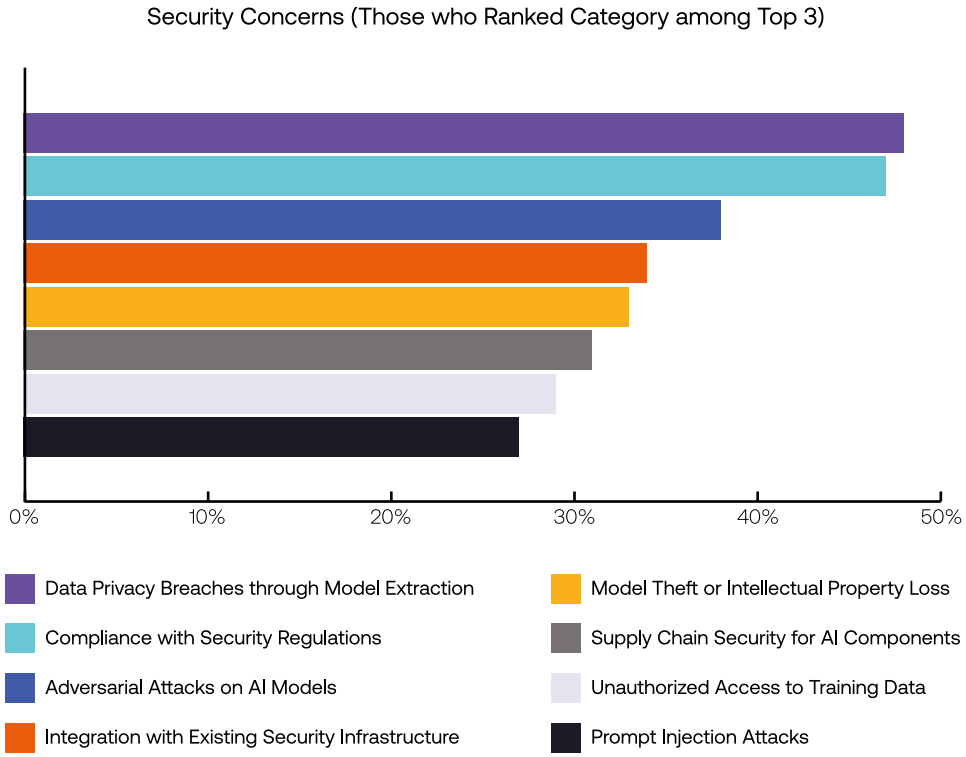
"We have our work cut out for us in formalizing these tool sets. Their use has largely grown organically within the business units, and that's why we have duplication."
- Vice President within the Information Technology department, working for an organization in the financial industry

Industry leaders recognize these challenges. One Vice President in the financial industry's IT department explained: "Our data quality and cataloging is so critical when it comes to AI...That is work in front of us." Another C-level IT executive in manufacturing emphasized the need for integrated data platforms: "We need to get the data into one platform so we can try to leverage the other functions to use AI."

These infrastructure, talent, and data challenges cannot be overlooked. Without addressing these fundamental capabilities, organizations risk their AI investments delivering diminishing returns as they attempt to scale beyond initial proof-of-concept deployments.

SECURITY SENSITIVITY

As organizations increasingly integrate AI into their operations, a new dimension of risk has emerged that demands urgent attention: data security and privacy. Security concerns are growing proportionally, with personally identifiable information (PII) rapidly becoming central to AI initiatives.



Nearly half (49%) of businesses already use customer data to power their AI initiatives, and the trend is accelerating. Looking ahead, 56% of leaders indicate their organizations plan to utilize personally identifiable information for future AI applications. This growing reliance on sensitive data introduces significant security considerations.

As one director in the life sciences and biotechnology industry explained: "I think when you get into some of these companies, ours in particular, we have to be a little bit more safeguarded because of the data sets we have.

Not all of it can be shared in public knowledge, but AI makes looking at that data, interpreting that data, analyzing that data 1000 times easier."

The expanded use of sensitive data in AI systems introduces security considerations and ethical concerns about bias and fairness. AI systems are only as objective as the data they're trained on, raising important questions about how organizations ensure their AI applications make fair and unbiased decisions.

"AI advancements will impact our organization by enhancing efficiency in decision-making and innovation across operations while requiring continuous adaptation to new technologies and regulatory changes."
- Business leader survey respondent

Our research reveals a significant gap in this area. Nearly half (47%) of business leaders acknowledge their organizations have limited bias detection and correction processes in their AI systems. Even more concerning, 17% reported having no formal bias correction process and relying instead on ad hoc bias checks. This inconsistent approach to bias mitigation poses significant risks, especially as AI becomes more deeply integrated into consequential business decisions.

Forward-thinking leaders recognize these challenges. As one Vice President in the financial industry observed: "We have to make sure that data is bias-free, and to do that, we have to come up with policies and standards controls that need to be implemented within the organization by support of data scientists and data stewards." This recognition drives 44% of leaders to identify AI ethics and data engineering as the most critical skills their organizations will need in the next five years.

On a more positive note, organizations are beginning to implement security measures for their AI systems. Only 5% of businesses report no specific AI security measures. The vast majority are taking at least some precautions, with 56% conducting regular security audits of AI models and infrastructure, 56% performing security testing of AI-powered applications, and 51% conducting vulnerability assessments for AI systems.

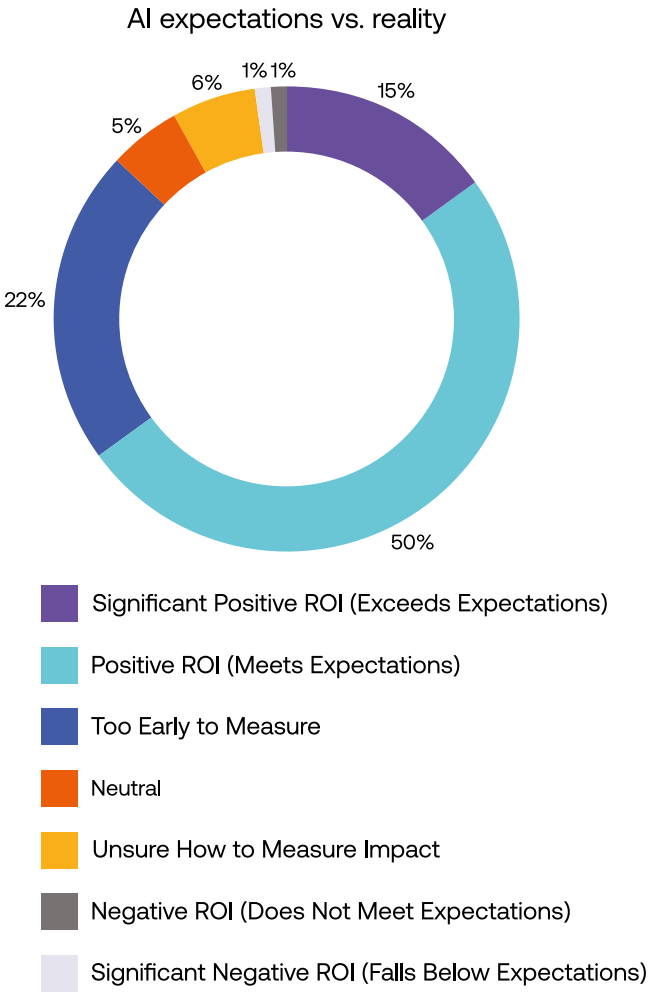
Despite these measures, significant concerns persist. Nearly half (48%) of business leaders identify data privacy breaches through model extraction as a top AI security concern. One Vice President of Marketing and Sales in retail observed: "I think there's a big open question on IT security that large companies are going to worry about."

As AI systems become more deeply integrated with sensitive business and customer data, organizations must evolve from basic security practices to comprehensive security strategies that address the unique vulnerabilities associated with AI technologies.

PREPARING FOR AN AI-POWERED FUTURE

One prediction seems certain as we look toward the horizon: AI adoption will continue to accelerate across all industries. Organizations that prepare effectively today will position themselves to capitalize on this technological revolution, while those that fail to address fundamental readiness gaps risk falling behind.

The business case for AI continues to strengthen among organizations using AI; 65% report that their AI applications are meeting or exceeding expectations for return on investment. This positive experience drives further adoption, with 92% of business leaders indicating plans to expand their AI usage in some capacity.



However, successfully scaling AI initiatives requires addressing several critical preparation gaps. A third (34%) of business leaders identify integration with existing security infrastructure as a top AI security concern. Meanwhile, 43% acknowledge their employees demonstrate only a basic understanding of AI at best, with additional training urgently needed. These challenges have led two-thirds (67%) of business leaders to identify AI security among the most critical AI-related skills their organizations will need to develop over the next five years.

Organizations must balance enthusiasm with strategic planning as they prepare for an AI-powered future. By systematically addressing

infrastructure limitations, closing talent gaps, enhancing data quality, and strengthening security measures, businesses can move beyond initial experimentation to realize AI's transformative potential at scale.

The path forward is clear, if challenging. Organizations must translate their AI enthusiasm into comprehensive strategies that address foundational readiness gaps. Those who successfully navigate this transition will be positioned to reap the efficiency gains and competitive advantages that AI promises. For the rest, the gap between AI ambition and AI readiness may prove increasingly difficult to bridge.

82%

of global business leaders are currently using AI applications.

39%

say their organization has a clearly defined, comprehensive AI strategy in place.

29%

of organizations can automatically scale compute or storage resources to meet AI demands.

23%

have dedicated power infrastructure to manage increased power demands for AI workloads.

95%

have implemented some kind of AI security measures.

arm

CHAPTER 2

The Technical Foundation: Technology Requirements for AI Success

DR JOHN SOLDATOS,
HONORARY RESEARCH FELLOW AT THE UNIVERSITY OF GLASGOW

2.1 OVERVIEW OF AI TECHNOLOGY REQUIREMENTS

2.1.1 The Need for a Robust Technology Infrastructure

Artificial Intelligence (AI) systems are advancing rapidly, becoming more flexible, scalable, and powerful. But to unlock their full potential, they need a solid technological foundation. Robust infrastructure is critical to managing the increasing complexity, scale, and sophistication of AI applications. It must support vast data processing, storage, and analysis while ensuring models run efficiently. As AI workloads grow, infrastructure must evolve to handle diverse tasks, reduce latency, and process massive data volumes—all while maintaining the speed and reliability modern applications demand.

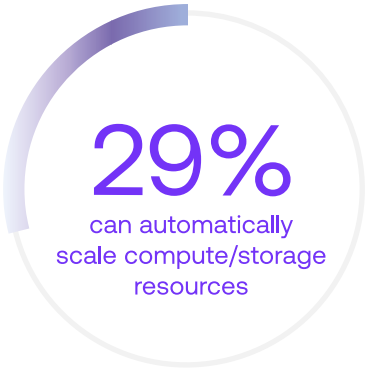
2.1.2 Technology Requirements: Main Challenges in Key Technological Areas

The development and deployment of robust AI technology infrastructures leverage various state-of-the-art developments, including high-performance computing architectures, power-efficient systems, scalable

infrastructures, and a proper balance between edge AI and cloud AI deployments. Recent advances in the above-listed areas help overcome considerable performance, scaling, and energy efficiency challenges. Specifically:

- **Evolution of computing architectures**: Over the past three decades we have witnessed a transition from simple, single-tier systems to complex multi-tier architectures that segment various aspects of an application (e.g., user interface, data access, business logic) into different functional layers. This transition has been driven by the need for more flexible, adaptable, and scalable systems. In line with this evolution, early AI systems were limited in their inability to efficiently process large datasets or perform complex computations. To address this limitation, there has been a shift towards multi-tier systems. This shift introduces multiple learning layers and service-oriented designs, enhancing system capabilities and performance.
- **Infrastructure scaling**: As AI models become larger and more sophisticated, there is a need for infrastructures that can scale in a cost-effective manner. As a prominent example, the size of Large Language Models (LLMs) and their training datasets are increasing at an unprecedented pace. This demands for infrastructures capable of scaling to accommodate both training data and data processing workloads. Recent advances in distributed computing offer solutions to these challenges, notably, solutions handle workloads at large scale without compromising speed or efficiency. Nevertheless, there is still a need for effective resource management strategies that optimize resource utilization while maintaining cost efficiency.
- **Edge vs. cloud computing considerations**: In recent years, organizations have managed workloads across the edge-cloud computing continuum. In this direction, the benefits of localized processing are confronted against those of centralized resources. In principle, edge computing reduces latency by processing data closer to its source, making it ideal for real-time applications. In contrast, cloud computing provides scalable

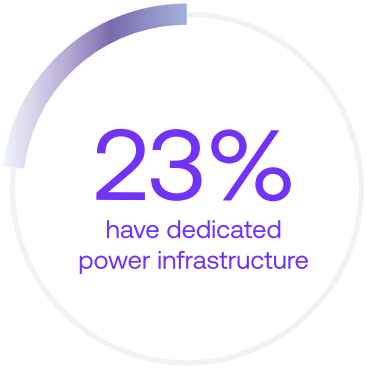
IT infrastructure readiness



access to virtually any number of computing resources, yet it also introduces latency issues for certain applications.

- **Power-performance trade-offs**: Today, energy consumption is one of the major concerns surrounding the deployment of AI systems at a very large scale. AI models are gradually growing in complexity, asking for more computational power and leads to increased energy consumption. Therefore, there is a need to balance high performance with energy efficiency, which is typically a significant challenge. At the same time, organizations must carefully consider the trade-offs between different processing architectures based on their specific workload requirements and constraints. Energy-efficient CPU architectures, like those based on Arm technology, have been crucial for enabling AI workloads at scale in data centers and edge computing environments. These processors combine general-purpose computing capabilities with optimizations for AI operations, providing an efficient foundation for inference workloads where versatility and power efficiency are paramount.
- **Specialized AI hardware**: Specialized AI hardware has also emerged to address specific computational challenges. Graphics Processing Units (GPUs) excel at parallel processing tasks common in AI training. Tensor Processing Units (TPUs) and Neural Processing Units (NPUs) are purpose-built for AI workloads, offering high performance per watt for specific types of neural network operations. Each of these technologies presents different trade-offs between computational power, energy efficiency, flexibility, and cost. Modern industrial enterprises are therefore offered a host of AI hardware options (e.g., AI-optimized CPUs, GPUs, TPUs, NPUs), which can be combined to meet their AI deployment requirements. In this context, companies must understand the capabilities of each option and how they can be integrated to serve their particular needs.

Power consumption management



2.1.3 User-driven technologies: Aligning infrastructure with end-user requirements

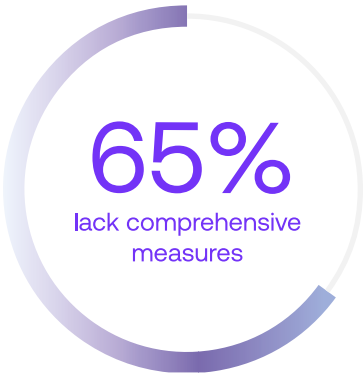
To ensure scalability, speed, and efficiency, it is also important to align technology infrastructure with end-user requirements. This involves understanding the specific needs of AI workloads, such as latency sensitivity and data volumes. In principle, AI infrastructures must provide tailored solutions for different AI application profiles in order to help ensure that performance, scalability, latency, energy efficiency, and other requirements are met. Emerging AI infrastructures must be driven by end-user requirements as much as they are driven by the evolution of AI technologies.

2.2 THE ARCHITECTURAL FOUNDATIONS OF MODERN AI

2.2.1 Edge-Cloud Integration: A New Paradigm for AI Infrastructure

The evolution of AI system architectures has played an essential role in enabling the development, deployment, and operation of state-of-the-art AI systems, such as edge-cloud AI systems and generative AI systems. For instance, large-scale generative AI systems (e.g., ChatGPT) and distributed learning systems (e.g., federated learning) could hardly be deployed based on old-fashioned, monolithic, single-tier architectures. Rather, they have become possible with the emergence of distributed, multi-tier, and hybrid architectures, which have opened new horizons for the scalability, adaptability, modularity, and performance of AI systems.

Energy efficiency optimization



In particular, AI architectures have undergone significant transformations over the decades. The main milestones in this evolution of AI architectures include:

- **Single-tier and multi-tiered systems**: As already outlined, AI systems were rarely monolithic. They consolidated user interface, logic, and data storage components into a single executable module. This offered

development and deployment simplicity yet limited scalability and adaptability. As such, single-tier architectures were unsuitable for non-trivial AI systems and tasks. To address these limitations, multi-tier (n-tier) architectures that segmented functionalities into specialized layers have emerged. Multi-tier systems enable modular development and distributed processing, which boost efficiency and scalability.

Cloud-based architectures: During the last two decades, the rise of cloud computing introduced on-demand access to computational resources, while also offering elasticity, capacity, and quality of services. The latter features have opened new horizons in the scalability and flexibility of enterprise workloads, including non-trivial AI applications.

- **Edge computing**: The advent of IoT devices, real-time applications, and mobile applications hosted in mobile devices like smartphones imposed new requirements that could hardly be met by cloud computing infrastructures. This led to the concept of edge computing, which is about processing data close to the data sources rather than to the cloud. Edge computing provides an effective solution for reducing latency, improving energy efficiency, and ensuring the confidentiality of sensitive datasets.

"Modern AI infrastructures typically employ a heterogeneous computing approach that leverages different processor types for different tasks."

Today, AI architectures are not purely cloud-based but instead support hybrid models. Hybrid architectures combine on-premises infrastructure with cloud resources to balance flexibility and control. For instance, in a hybrid deployment, on-premises systems can be used to handle sensitive data and/or latency-critical tasks, while cloud resources are used to provide scalability for training large models or storing vast datasets. Advanced hybrid solutions integrate cloud and edge computing to optimize performance. For instance, edge devices preprocess data locally before sending it to the cloud for deeper analysis. This helps leverage the advantage of both edge and cloud. In particular, edge-cloud deployments combine the scalability, cost flexibility (pay-as-you-go), and intelligent centralized resource management capabilities of the cloud with the reduced latency and enhanced privacy characteristics of the edge.

2.2.2 AI Models and Hardware

The evolution of AI system architectures has played an essential role in enabling the development and deployment of different AI models. Specifically, modern AI systems integrate a variety of models that feature different architectures, including:

- **Neural networks**: Neural networks are the foundational architecture for many AI systems. They consist of interconnected nodes (neurons) that mimic the human brain's structure. Prominent types of Neural Networks include

  (i) **Feedforward Neural Networks (FNNs)**, where data flows in one direction, from input to output. FNNs are used in basic classification and regression tasks;
- (ii) **Convolutional Neural Networks (CNNs)**, which are specialized for image processing and pattern recognition. As a prominent example, the AlexNet CNN, revolutionized image classification by using deep convolutional layers;
- (iii) **Recurrent Neural Networks (RNNs)**, which are designed for sequential data with feedback loops to retain memory. As a prominent example, Long Short-Term Memory (LSTM) networks are extensively used in time-series forecasting and natural language processing (NLP);
- (iv) **Residual Networks (ResNet)**, which introduced skip connections to address vanishing gradient issues in deep networks. As an example, ResNet50 is widely used in computer vision tasks like object detection. Overall, Neural networks are used in a wide variety of applications and tasks, including image recognition, speech-to-text systems, medical diagnosis, and more.
- **Transformer models**: Transformers have revolutionized AI by enabling parallel processing and self-attention mechanisms in order to handle sequential data in very efficient ways. From an architecture viewpoint, a transformer model consists of an encoder-decoder structure. The encoder processes input sequences into representations. At the same time, the decoder generates outputs based on these representations.

The core components of transformers also include multi-head attention layers and feedforward networks. One of the most prominent examples of transformer models is **BERT (Bidirectional Encoder Representations from Transformers)**, which is an encoder-only model for tasks like text classification and sentiment analysis. Another example is the **T5 (Text-to-Text Transfer Transformer)** model, which is an encoder-decoder model for translation, summarization, and other NLP tasks. Most importantly, the very popular GPT (Generative Pre-trained Transformer) a family of models, i.e., the models behind ChatGPT, are also decoder-only models for text generation and conversational AI. In general, transformers excel in NLP tasks such as machine translation, text summarization, and question answering. They are also applied in protein folding (e.g., AlphaFold) and computer vision tasks like object detection.

- **Generative AI Models**: Generative models create new data resembling the input data they were trained on. As a prominent example, Generative Adversarial Networks (GANs) comprise a generator that creates synthetic data and a discriminator that evaluates its authenticity. Notable examples of GANs are the DCGAN (Deep Convolutional Generative Adversarial Networks) for image generation and the CycleGAN for domain transfer, like turning photos into paintings. Another class of Generative AI Models is the Variational Autoencoders (VAEs), which learn latent representations to generate new data points. VAEs are used to generate realistic images or interpolate between data points. Overall, generative models are used in creative tasks like art generation, synthetic data creation, and drug discovery.
- **Multi-modal models**: These models process and integrate multiple types of data (e.g., text, images). For instance, **CLIP (Contrastive Language–Image Pre-training)** aligns visual inputs with textual descriptions for tasks like image captioning or zero-shot classification. As another example, **DALL-E**: Generates images based on textual prompts by combining NLP and computer vision capabilities.

Multi-modal models are usually applied in content creation, augmented reality, and cross-modal retrieval systems.

- **Other specialized architectures**: There are also other models tailored for specific domains or tasks. For instance, **Vision Transformers (ViTs)** adapt transformers for image processing by dividing images into patches. They are used in tasks like object detection and segmentation. As another example, **Deep Reinforcement Learning Models** combine neural networks with reinforcement learning techniques. The AlphaZero model, notorious for mastering board games like chess and GO, is a deep reinforcement learning model.

Overall, AI system architectures have evolved from basic neural networks to advanced designs like transformers and generative models. Each category serves a unique purpose. Neural networks provide foundational AI capabilities, while transformers dominate NLP and multi-modal applications. Generative AI models enable creative outputs, and multi-modal models bridge different data types.

Moreover, modern AI architectures and the above-listed AI models benefit from specialized hardware. Specifically:

- **CPUs** are fundamental to AI infrastructure, particularly for inference workloads where their versatility and power efficiency are crucial. Modern CPU architectures optimized for AI workloads combine sophisticated vector processing capabilities with energy efficiency, making them ideal for deploying AI at scale in datacenters and edge devices. Their ability to efficiently handle diverse workloads, from preprocessing to inference, makes CPUs essential for real-world AI deployments where flexibility and cost-effectiveness are paramount.
- **GPUs** excel at parallel processing. Their ability to handle matrix operations efficiently make them indispensable for training deep learning models. Today, AI companies are massively deploying GPUs to enable the provision of their products or services at scale.

– **NPUs** are purpose-built processors designed specifically for accelerating neural network computations. By incorporating dedicated circuitry for common AI operations like convolutions and matrix multiplication, NPUs achieve high performance per watt for both training and inference tasks. Their architecture makes them particularly effective for specialized edge AI applications where power constraints are critical, enabling sophisticated AI capabilities in mobile devices, IoT sensors, and other resource-constrained environments.
– **TPUs** are optimized for tensor-based operations, which provide high throughput for large-scale deep-learning tasks. While GPUs are versatile across various applications, TPUs tend to specialize in accelerating neural network computations.

AI-optimized hardware solutions are significantly reducing training and inferencing times and computational costs for complex AI models.

2.2.3 Trends in AI Systems and Architectures

Some of the most prominent AI systems trends and their AI hardware implications include:

– **Integration with machine learning frameworks**: Nowadays, distributed computing platforms (e.g., Apache Spark or Hadoop) are increasingly integrated with frameworks such as TensorFlow and PyTorch. This combination enhances scalability while simplifying model deployment across distributed environments.
– **Modular datacenters**: Modular designs allow datacenters to scale incrementally by adding or replacing components as needed. This approach caters to increased scalability and sustainability, as it reduces both energy consumption and construction costs.
– **Distributed training of AI models**: Distributed training techniques take advantage of multiple computing nodes to enable faster model training. This is particularly beneficial for large-scale transformer models like GPT

(Generative Pre-training Transformer) or BERT (Bidirectional Encoder Representations from Transformers).

- **Federated learning**: Federated learning allows decentralized devices to collaboratively train global, more accurate models without sharing raw data. This approach enhances privacy while reducing bandwidth usage.
- **Fine-tuning**: Fine-tuning adapts a pre-trained model to perform specialized tasks with greater accuracy. Instead of training a model from scratch, fine-tuning leverages the knowledge embedded in a pre-trained model, such as GPT or BERT, and refines it using task-specific data. This approach reduces computational costs and training time while improving performance on domain-specific tasks. Fine-tuning can range from full optimization of all model layers to partial fine-tuning, where only specific layers or parameters are updated. It is already widely used for tasks such as customizing conversational tones in chatbots, enhancing domain-specific knowledge in legal or medical applications, and addressing edge cases not covered during pre-training.
- **Retrieval-Augmented Generation (RAG)**: RAG enhances the capabilities of generative models by integrating real-time information retrieval. Unlike traditional large language models (LLMs) that rely solely on static training data, RAG incorporates external data sources (e.g., databases, APIs, document repositories) into its responses. This dynamic retrieval process allows RAG systems to provide more accurate, context-aware, and up-to-date outputs. The process involves indexing relevant data into vector databases, retrieving pertinent information based on user queries, augmenting the input with this data, and generating responses that combine both retrieved and pre-trained knowledge. RAG is already used in numerous applications, including advanced question-answering systems, content summarization, conversational agents, and personalized educational tools. Emerging trends in RAG include hybrid search techniques that combine structured and unstructured data retrieval, multimodal RAG for integrating text with images or audio, and on-device RAG for enhanced privacy and reduced latency.

"Recent studies by the International Energy Agency estimate that AI workloads now account for 10-15% of total electricity usage in data centers, with projections suggesting this could reach 25-30% by 2030."

RAG demands efficient data processing and inference based on the following key hardware components: (i) GPUs (e.g. NVIDIA Hopper and Blackwell GPUs), AI-optimized CPUs (e.g. Arm Neoverse-based CPUs in the cloud, such as Google Axion Processors) or Heterogeneous compute solutions combining CPUs and GPUs (e.g. Arm Neoverse-powered NVIDIA Grace Hopper) for embedding generating and LLM inference. These help ensure high throughput and low latency during real-time retrieval and response generation. (ii) GPUs for embedding generation and LLM inference. These help ensure high throughput and low latency during real-time retrieval and response generation; (iii) Random Access Memory (RAM) (e.g., 64GB or more) to properly handle embeddings and vector databases; (iv) Fast NVMe SSDs for storing large datasets and vector indices in ways that help ensure quick access during retrieval steps; (v) Scalability solutions such as cloud-based solutions that can offload storage and search workloads in order to reduce the need for extensive on-premises hardware while maintaining scalability.

- **Agentic AI**: Agentic AI refers to AI systems designed to operate autonomously while exhibiting decision-making capabilities similar to human agents. These systems go beyond passive data processing to actively interact with their environment, adapt based on feedback, and pursue goals independently. Agentic AI incorporates elements like reinforcement learning and multi-agent collaboration to optimize its performance in dynamic settings. For instance, agentic AI systems can autonomously navigate supply chain logistics or manage financial portfolios based on real-time data. Agentic AI is closely tied to advancements in areas like reinforcement, fine-tuning and adaptive learning. These techniques enable AI agents to refine their decision-making processes over time while maintaining alignment with predefined objectives. Agentic AI systems are complex as they involve autonomous decision-making, learning, and action modules. Their hardware demands are driven by the need to process multimodal data, execute actions in real time, and adapt dynamically.

(i) AI-optimized, power efficient CPUs (such as Arm Neoverse-based CPUs) paired with GPUs like NVIDIA Hopper or Blackwell to support the perception module (such as computer vision) and decision-making engines that rely on reinforcement learning;

(ii) **Large memory capacities (128GB RAM or more)** are needed to manage simultaneous tasks across multiple agents in a modular architecture;

(iii) **NVMe SSDs** ensure fast access to knowledge graphs, training datasets, and historical data used by cognitive modules;

(iv) Networking & Power as distributed Agentic AI systems require robust networking infrastructure to facilitate communication between agents. Power-efficient designs are, therefore, critical to supporting long-term operations in autonomous environments.

2.3 POWER AND PERFORMANCE SOLUTIONS

2.3.1 Computational Demands

Large-scale AI models, such as large language models (LLMs) and generative AI systems, are growing exponentially in size and complexity. These models consist of hundreds of billions or even trillions of parameters, requiring large amounts of computational resources for efficient training and deployment.

Modern AI infrastructures typically employ a heterogeneous computing approach that leverages different processor types for different tasks. The foundation often begins with high-performance CPUs, which provide the versatile computing capabilities needed for data preprocessing, inference, and orchestrating complex AI workloads. Energy-efficient CPU architectures have proven particularly valuable for inference at scale, where power efficiency and consistent performance are crucial.

For training-intensive workloads, organizations often augment their CPU infrastructure with specialized hardware accelerators. GPUs excel at the parallel processing needed for model training, while NPUs offer optimized performance for specific AI operations. This combination of general-purpose processors and specialized accelerators helps organizations balance performance, efficiency, and cost-effectiveness across their AI workflows.

The sheer scale of computational requirements continues to push the limits of existing infrastructures. Training state of the art LLMs like GPT-4 or Meta Llama 3 involves processing vast datasets using high-performance computing (HPC) setups that typically combine multiple types of processors working in concert. In this context, distributed computing frameworks and advanced hardware architectures that can effectively coordinate these various processing elements are required to meet these growing demands.

2.3.2 Computational Demands for Training and Inferencing

AI models have considerable computational demands at both their training and inferencing stages. These demands are significant given the complexity of models, the size of datasets, and the need for real-time performance. The hardware options previously presented play a considerable role in meeting these demands, providing specialized components optimized for different stages of AI workflows.

Hardware Support for AI Training

Training AI models is highly resource-intensive. It involves massive floating-point operations (FLOPs). For instance, training LLMs like GPT-4 requires tens of thousands of GPUs and consumes tens of gigawatt-hours of energy. Key factors influencing computational demand include:

- **Model complexity**, as larger models with billions or trillions of parameters require more compute power.
- **Dataset size,** as larger datasets increase training time and resource needs.

– **Parallel processing**, given that the efficient use of multiple processors is important when it comes to handling large-scale computations.

In this context, modern hardware for training AI includes:

– GPUs like the NVIDIA A100 or NVIDIA GeForce RTX™ 4090 are widely used due to their high parallel processing capabilities and ability to deliver trillions of operations per second (TOPS).
– Google's TPUs are optimized for deep learning tasks while offering high performance in datacenters.
– ASICs (Application-Specific Integrated Circuits), i.e., custom chips designed for specific AI tasks that provide energy efficiency and speed.
– Compute solutions combining CPU cores and GPUs like NVIDIA's Arm-based Grace Blackwell and Grace Hopper offer significant uplifts in AI training performance.

Hardware Support for AI Inferencing

Inferencing involves the application of trained models to new data in order to make predictions or decisions. While less computationally intense than training, they demand low latency and high throughput. The latter are very important in cases of real-time applications like autonomous vehicles or healthcare diagnostics. In this context, inferencing hardware focuses on speed and efficiency, including:

– Compact and power-efficient devices like NVIDIA Jetson Nano™ enable real-time inferencing at the edge.
– NPUs are typically specialized for inferencing tasks and can balance performance with energy efficiency.
– Solutions combining GPUs and CPUs (e.g. NVIDIA Grace Hopper superchips) are great choices for high-throughput, real-time inferencing of large models.
– CPUs (e.g. Arm-based AWS Graviton processors) are alternatives for

many inference scenarios, offering strong price-performance and energy efficiency—especially for smaller models.

There are also a variety of optimization techniques that are aimed to enhance inferencing performance, including:

- **Quantization**, which involves reducing model precision (e.g., from 32-bit to 8-bit) in order to improve speed and reduce memory usage.
- **Pruning**, which involves removing redundant model parameters to lower computational complexity without sacrificing accuracy.
- **Batching**, which processes multiple inputs simultaneously towards maximum GPU utilization.

The future of AI training and inferencing will be shaped by emerging trends in hardware development, including:

- **Distributed computing**, as inferencing is increasingly moving toward distributed systems to handle growing demands.
- **Energy efficiency**, given that advances in memory architectures and specialized chips aim to reduce energy consumption during both training and inferencing.
- **Real-time applications** based on edge computing which is becoming important for applications that require immediate responses (e.g., autonomous driving).

2.3.3 Energy Efficiency

AI workloads are notoriously energy-intensive, leveraging datacenters that consume significant power for computation and cooling. It is estimated that AI workloads account for 10–20 percent of total electricity usage in datacenters. This demand raises sustainability concerns due to the associated carbon emissions. Overall, the main AI-related energy efficiency challenges are:

- **Datacenter power consumption**: The electricity required to train large AI models can be staggering. For instance, training a single LLM can consume energy equivalent to powering hundreds of homes for a year. To make matters worse, cooling systems are also highly energy-intensive.
- **Carbon footprint**: The environmental impact of AI is substantial and represents one of the main setbacks to its wider adoption. In particular, according to the International Energy Agency, modern datacenters contribute approximately 1 percent of global greenhouse gas emissions.

"Training a single LLM can produce carbon emissions equivalent to five cars over their lifetimes."

To address these power consumption challenges, AI vendors and integrators leverage the following solutions:

- **Energy-efficient chips**: Recent innovations in chip design are mitigating energy efficiency challenges. For example, the low-power Arm architecture optimizes performance per watt by minimizing energy consumption without sacrificing computational capability.
- **Sustainable datacenters**: Modular datacenters powered by renewable energy sources are emerging as a solution to reduce the carbon footprint of AI infrastructures. Furthermore, the use of advanced cooling technologies and virtualization techniques also improves operational efficiency.

2.3.4 Latency Sensitivity

Latency is another critical factor in AI applications, especially for the class of AI applications that incorporate real-time decision-making, such as autonomous vehicles, healthcare monitoring, and industrial automation

applications. For instance, autonomous vehicles must process sensor data in milliseconds to make driving decisions, while wearable healthcare devices need low latency to monitor vital signs and alert users in near real-time.

To address latency challenges, AI vendors and integrators are employing local data processing at the edge of the AI deployment. Edge computing processes data closer to its source, which reduces round-trip times and improves responsiveness. The proliferation of AI applications with real-time constraints has given rise to the edge AI paradigm, which deploys powerful AI algorithms (e.g., deep neural networks) on edge devices like gateways or edge servers. One of the most prominent edge AI approaches is the TinyML paradigm, which deploys smaller AI models on ultra-low-power devices like microcontrollers towards enabling lightweight machine learning tasks with minimal latency. Edge AI systems commonly integrate hardware-level optimizations, which leverage advancements in memory bandwidth and dedicated AI accelerators to help minimize latency during AI computations. Apart from boosting real-time responsiveness, edge AI systems enhance privacy and improve bandwidth efficiency, as significantly less data is transmitted over the networking part of the AI infrastructure.

2.4 INFRASTRUCTURE SCALING REQUIREMENTS

2.4.1 Scalability Needs

AI infrastructures must scale in a cost-effective fashion in order to cope with growing data volumes and the increasing complexity of AI models. Large-scale AI workloads, such as training large language models (LLMs) and real-time inference systems, demand immense computational power and efficient resources. The exponential increase in data and compute demands leads to bottlenecks arising in storage, processing speed, and network bandwidth. Moreover, maintaining system performance while scaling infrastructure can be costly and operationally complex. To the rescue, distributed computing offers a scalable approach by dividing

workloads across multiple nodes or machines. Parallel processing is also key for handling large-scale AI tasks more efficiently. In this direction, open-source platforms like Kubernetes enable the orchestration of distributed resources toward optimizing training times and inference processes.

2.4.2 Cost Management

Balancing performance with cost-effectiveness is another consideration when it comes to scaling AI infrastructures. Modern enterprises must navigate trade-offs between cloud-based elasticity and on-premises control to optimize resource utilization. Cloud computing provides dynamic scalability, which allows organizations to adjust resources based on workload demands without significant upfront investment. However, this flexibility comes at the cost of reduced control over infrastructure and potential long-term expenses for high usage. On the other hand, on-premises solutions offer greater control and customization but require substantial capital investments and ongoing maintenance costs.

At the same time, resource pooling is an effective strategy for optimizing computing utilization. Based on the aggregation of resources across multiple users or regions, organizations can maximize efficiency while minimizing idle time. For example, pooling workloads with non-overlapping peak demands allows for better utilization of shared resources. This approach reduces the need for over-provisioning while ensuring computational capacity during peak periods.

2.4.3 Modular Approaches

Modern modular infrastructure designs provide flexible solutions for scaling AI systems in an incremental fashion without requiring complete overhauls. Specifically, modular datacenters allow organizations to add or upgrade components as needed, enabling incremental scaling to meet growing demands. Modular datacenters are particularly valuable for edge computing environments where there are considerable space and power constraints.

2.5 EMERGING TRENDS IN AI INFRASTRUCTURE

2.5.1 Sustainability Initiatives: Shift Towards Greener Technologies

In an era where environmental performance and sustainability are at the very top of the political and enterprise agendas, there is a shift towards greener technologies to reduce the carbon footprint associated with AI infrastructure. This shift is powered by the following trends:

- **Renewable energy sources and green datacentres**: AI datacenters are transitioning to renewable energy sources such as solar and wind power. This shift not only reduces greenhouse gas emissions but also aligns with global sustainability goals. Companies like Google and Microsoft have committed to running their datacenters on 100% renewable energy, which set the bar high for other infrastructure providers in the industry. Moreover, Google has recently partnered with Kairos Power to deploy compact nuclear reactors for its AI-driven data centers. At the same time, Microsoft is funding the restart of a mothballed reactor at Three Mile Island to supply power for its operations by 2028.
- **Energy-efficient chips**: As already outlined, chip design innovations contribute to reducing energy consumption. Arm, for example, has developed low-power architectures that optimize performance per watt. These chips are designed to handle complex AI tasks while consuming less energy, which makes them ideal for both datacenters and edge devices.

2.5.2 Infrastructure Management Automation

Automation is becoming increasingly important in managing AI infrastructures. Generative AI (GenAI) tools have recently had a prominent role in realizing this automation. These tools enhance resource provisioning, workload optimization, and scaling decisions, which improve operational efficiency automation. More specifically, during the last couple of years, GenAI technologies have revolutionized resource allocation through auto-scaling features. Tools like K8sGPT enable GenAI-powered resource

management that adjusts resource provisioning dynamically based on real-time workload requirements. This helps to ensure that computational resources are allocated efficiently during peak usage while scaling down during low-demand periods. This level of automation reduces the need for manual intervention, which allows IT teams to focus on strategic initiatives rather than routine maintenance tasks. Overall, automation reduces operational costs without compromising performance.

Conclusion

"Edge computing processes data closer to its source, which reduces round-trip times and improves responsiveness."

As AI applications grow in size and complexity, there is a clear need for evolving the AI technology infrastructures. For over two decades, we have witnessed significant improvements in different aspects of AI infrastructures, including hardware, software, and architecture-related aspects. For instance, the transition from monolithic single-tier systems to distributed, multi-tier, and hybrid architectures has played an important role in addressing the scalability, adaptability, and performance challenges of AI systems. Nowadays, modern architectures integrate cutting-edge technologies such as CPUs, GPUs, TPUs, and microservices-based designs to enhance computational efficiency and flexibility.

Key infrastructure-related challenges such as balancing power-performance trade-offs, managing large-scale data processing, and minimizing latency have been effectively addressed through innovations in hardware and infrastructure design. In recent years, edge-cloud computing paradigms have emerged as a cornerstone for achieving low-latency processing while maintaining scalability and privacy. Furthermore, sustainability has taken center stage based on initiatives that reduce the environmental impact of AI operations. Prominent examples of such initiatives include renewable energy-powered datacenters and energy-efficient chip designs.

Moreover, the latest advances in automation are transforming the infrastructure management landscape. In the coming years, GenAI tools

that dynamically optimize resource allocation are expected to reduce operational costs while maintaining peak performance. Also, modular approaches to scaling infrastructure will provide flexibility for incremental growth without overhauling existing systems.

The above-listed advancements will enable businesses to overcome technical barriers and unlock the full potential of AI-driven solutions. Companies like Arm are playing a significant role in this transformation through innovative chip designs and ecosystem partnerships that address energy efficiency and computational demands. The current state and outlook of the AI technology infrastructure indicate that the future of AI will include not only technological breakthroughs but also a more efficient and environmentally responsible digital ecosystem.

ARM SIDEBAR

AI Everywhere: Arm's Vision for the Future of Compute

By Kevork Kechichian, EVP, Solutions Engineering, Arm

Artificial intelligence (AI) is reshaping computing, influencing everything from cloud-scale processing to real-time inference on personal devices. As AI adoption accelerates, the demand for high-performance, scalable, and energy-efficient compute platforms continues to grow. The challenge lies in enabling AI across a diverse range of environments—from cloud datacenters to power-constrained edge devices—without compromising efficiency or security.

The Arm compute platform plays a critical role in this transformation, providing the architectural foundation for AI workloads across industries. By enabling AI to run efficiently on CPUs, GPUs, and NPUs, Arm supports an ecosystem where AI can be deployed at scale. The approach is built on three fundamental pillars: accelerating AI from cloud to edge, delivering energy-efficient AI inference, and ensuring security and trust in AI workloads.

Accelerated AI Everywhere #onArm From Cloud to Edge

AI workloads span a vast spectrum of compute environments, ranging from cloud-based model training to on-device inference. The ability to run AI efficiently across this continuum is crucial for scaling AI applications and ensuring real-time performance. The Arm compute architecture is designed to support AI workloads at every level, enabling accelerated AI from cloud to edge.

In cloud environments, Arm-powered solutions (Nasdaq:ARM) provide scalable AI compute, helping datacenters optimize performance while maintaining energy efficiency. Arm CPUs are increasingly being adopted in cloud instances for AI inference, as they deliver high computational throughput at lower power consumption compared to traditional architectures. As AI models grow in complexity, cloud providers require efficient compute platforms that can handle large-scale workloads while managing operational costs.

Beyond the cloud, AI inference is rapidly shifting toward the edge, driven by the need for low-latency processing and enhanced privacy. Running AI on-device—whether on smartphones, industrial sensors, or embedded systems—reduces dependency on cloud resources and enables real-time decision-making. The Arm processor architecture facilitates this shift by providing a flexible foundation that supports diverse AI workloads, ensuring optimal performance across different hardware configurations.

The Armv9 architecture introduces key enhancements for AI acceleration, including Scalable Vector Extensions (SVE), which improve AI compute efficiency, and the Compute Matrix Engine (CME), which enhances vector processing capabilities. Additionally, Arm's heterogeneous computing

approach allows CPUs to work alongside GPUs and NPUs, providing a balanced AI execution framework tailored to specific workload demands.

As AI continues to scale, the ability to run workloads seamlessly from cloud to edge will become increasingly important. The Arm compute platform helps ensure that AI applications can be deployed across a broad range of devices while maintaining efficiency, flexibility, and scalability.

The Energy-Efficient Pervasive Compute Platform to Power GenAI

The exponential growth of AI workloads has raised concerns about power consumption and sustainability. Training large AI models requires significant computational resources, while AI inference at scale demands efficient processing to avoid excessive energy usage. The Arm compute architecture is designed to address this challenge, enabling AI workloads to run with high efficiency without compromising performance.

The power-efficient design philosophy of Arm is rooted in decades of innovation. The Armv9 architecture integrates advanced features that optimize AI compute while minimizing energy consumption. Technologies such as Helium for machine learning acceleration in IoT devices and SVE for improved AI inference performance allow AI applications to run with reduced power overhead. These innovations ensure that AI workloads can scale efficiently, whether in datacenters, edge devices, or battery-operated consumer electronics.

One of the key advantages of Arm-based solutions is their ability to provide high-performance AI inference while maintaining low power requirements. In the mobile industry, for example, over 70 percent of AI workloads in third-party applications already run on Arm CPUs. This trend is expanding

into PCs, industrial automation, and autonomous systems, where AI-driven applications require continuous processing within strict power constraints.

Sustainability is becoming a central focus in AI deployment, particularly in large-scale cloud operations. AI workloads contribute to increasing power demands in datacenters. By leveraging Arm-based compute platforms, cloud providers can optimize energy efficiency, reducing power consumption while maintaining high compute performance.

The Developer-Trusted Platform for Running AI Workloads Everywhere

As AI adoption grows, security has become a critical factor in ensuring trust and reliability in AI workloads. AI models process vast amounts of data, often handling sensitive or proprietary information, making protection against security threats, data breaches, and unauthorized access essential.

"Arm's compute platform plays a critical role in this transformation, providing the architectural foundation for AI workloads across industries."

Arm takes a security-first approach, integrating robust security features into its compute architectures to provide a trusted foundation for AI deployment across industries. The Armv9 architecture introduces Arm Confidential Compute Architecture (CCA) and Realm Management Extension (RME), ensuring that silicon protects AI data and code wherever computing happens. These security enhancements are crucial as AI inference shifts from cloud to edge, requiring devices to process sensitive data locally without exposing it to external threats.

By integrating hardware-based security features, the Arm architecture enables privacy-preserving AI, safeguarding data at rest and in transit. This is especially relevant in sectors such as healthcare, finance, and autonomous systems, where AI models handle confidential information. Additionally, in cloud environments, the Arm security framework provides

hardware-enforced protections that mitigate risks associated with multi-tenant infrastructures and prevent unauthorized model access.

As AI continues to scale, security must remain at the core of compute design. By embedding security at the architectural level, Arm ensures that AI applications operate in a trusted, resilient, and privacy-focused environment, empowering organizations to deploy AI workloads with confidence.

CHAPTER 3

Policy and Governance: Shaping the AI Regulatory Landscape

DR NORA VON INGERSLEBEN SEIP, POSTDOCTORAL RESEARCHER AT THE POLITICAL SCIENCE DEPARTMENT OF THE UNIVERSITY OF AMSTERDAM, REGULAITE PROJECT

The policy and governance landscape for artificial intelligence (AI) is multifaceted and complex due to several factors. First, a variety of actors at local, national, regional, and global levels contribute to the governance of AI. Second, the umbrella term "AI" itself spans diverse technologies, from robots to large language models (LLMs), necessitating a broad spectrum of regulations that impact AI across its lifecycle—design, development, and deployment.

3.1 UNDERSTANDING THE POLICY AND REGULATORY PILLAR OF AI

During the design and development phases, concerns about data provenance, bias, and transparency are predominant. The deployment phase focuses on preventing unethical uses, such as AI, in mass surveillance systems targeting political dissidents. This results in an intricate array of rules by various authorities governing AI's lifecycle, creating complexity but also enhancing predictability while ensuring safety and accountability.

Regulatory preparation

To get a better understanding of the governance challenges posed by AI, AI can be categorized into foundation models, AI-powered physical products, small-scale AI as a service, and militarily relevant AI, each presenting unique risks and requiring specific governance strategies. Foundation models (large-scale AI models trained on vast amounts of diverse data, capable of performing a wide range of tasks)known for their scale and versatility, pose significant challenges, such as the potential misuse by malicious actors and complications arising from their evolution towards multi-modal capacities that inch towards artificial general intelligence. Safety summits and proactive regulations are crucial in addressing these concerns.

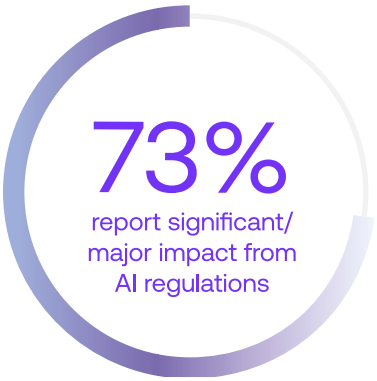
AI also permeates physical products, from everyday items like appliances to critical devices such as medical equipment, necessitating robust safety standards and cyber resilience. The distinction between high-risk and low-risk AI applications helps in tailoring governance appropriately. For digital AI services, which easily cross jurisdictions, international cooperation is essential to manage risks related to user safety and privacy.

In the military sphere, AI enhances capabilities from logistics to combat operations, raising ethical issues regarding human oversight and conflict escalation. The diverse applications of AI across these categories underscore the need for comprehensive and adaptable governance frameworks to mitigate risks while fostering innovation.

To address these diverse governance challenges, countries and regions have developed varying regulatory frameworks tailored to their specific priorities and concerns regarding AI development and deployment.

3.2 CURRENT REGULATORY FRAMEWORKS

Impact of regulations



While there are many private and multi-stakeholder initiatives formulating AI policy and governance frameworks, only governments and the European Commission have the power to make general laws and regulations that are automatically and directly binding on their populations. Below is an overview of the most important governmental initiatives in the realm of AI policy and governance.

To address these diverse governance challenges, countries and regions have developed varying regulatory frameworks tailored to their specific priorities and concerns regarding AI development and deployment.

3.2.1 Overview of Existing Frameworks

In August 2024, the European Union signed the AI Act, the world's first comprehensive legal framework for AI. The AI Act uses a risk-based approach, meaning that AI applications are regulated differently depending on whether they pose an "unacceptable risk," a "high risk," or a "limited risk." Applications that pose a clear threat to safety, fundamental rights, or democratic values fall into the "unacceptable risk" category and are strictly prohibited under the AI Act. Examples are social scoring by governments, real-time biometric identification in public spaces, and predictive policing systems. Applications that have a significant impact on people's safety, rights, or livelihoods fall into the "high-risk" category. These systems are not forbidden but are subject to stringent transparency requirements and need to go through conformity assessments to ensure their responsible use.

"By integrating hardware-based security features, Arm's architecture enables privacy-preserving AI, safeguarding data at rest and in transit."

Examples are AI systems that manage critical infrastructure, determine access to education, or are used in recruitment processes. AI applications that are thought to not have the same impact on safety or fundamental rights as high-risk systems, e.g., chatbots, recommender systems, and video editing tools, fall into the "limited risk" category. These applications

require compliance with basic transparency obligations to ensure their ethical and transparent use. The AI Act empowers regulatory bodies, led by the newly created EU AI Office, to monitor compliance with the Act and enforce regulations, with steep fines for violations.

"I think when you get into some of these companies, ours in particular, we have to be a little bit more safeguarded because of the data sets we have, right? Not all of it can be shared in public knowledge." *- Director of Operations, working for the Life-Sciences/ Biotechnology industry*

In contrast to the EU, the United States currently does not have comprehensive, federal AI-specific regulation akin to the EU AI Act. Instead, AI applications are primarily regulated through existing laws and sector-specific regulations in areas such as healthcare, finance, transportation, and employment. Additionally, there are state-level laws, e.g., the California Consumer Privacy Act and the Illinois Biometric Information Privacy Act, that indirectly regulate AI systems and applications. At the federal level, the National AI Initiative Act of 2020 aims to advance AI research, development, and use in a coordinated and strategic manner across the country to ensure that the United States remains a global leader in AI. Moreover, federal agencies such as the National Institute of Standards and Technology (NIST) and the Federal Trade Commission have released guidelines for assessing and mitigating AI risks. There are also multiple voluntary ethical guidelines adopted by major tech companies such as Microsoft, Google, and OpenAI. (For a more detailed analysis of the evolving US regulatory landscape and its sector-specific approach, see the following chapter which provides insights on recent policy shifts and their implications for AI developers.)

In Asia Pacific, China and Singapore stand out as the two countries that have been most active in regulating AI. Policy and governance initiatives in China are led by the central government and are aimed at maintaining rapid technological development while ensuring state control and alignment with the Chinese Communist Party's political and social priorities. Key pillars of China's AI governance framework include the New Generation Artificial Intelligence Development Plan, the Internet Information Service Algorithmic Regulation Provisions, the Data Security Law and

"I think there's a big open question on IT security that large companies are going to worry about."
- Vice President of Marketing/Sales, working in retail industry

Personal Information Protection Law, and the Ethical Guidelines for AI. China has also enacted targeted regulations for specific high-impact AI applications such as facial recognition, deepfakes, and autonomous vehicles. Overall, China's AI governance landscape reflects the country's ambition to lead globally in AI development while maintaining strict control over AI's social and political implications.

Singapore's approach to AI governance is guided by the Model AI Governance Framework, first introduced in 2019 and updated in 2020 by the Infocomm Media Development Authority and the Personal Data Protection Commission. Key principles meant to be advanced through the Model Framework are human-centricity, transparency, fairness, and accountability. The framework provides practical guidelines for businesses to deploy AI responsibly, focused on ensuring explainability, robustness, and stakeholder involvement. Singapore encourages organizations to conduct self-assessments using the Model Framework and the AI Verify toolkit, introduced in 2022. Overall, Singapore aims to balance regulation with growth and to be a global leader in trusted AI adoption.

Other notable AI policy and governance initiatives include Canada's AI and Data Act (AIDA) and Brazil's draft AI bill. Canada's AIDA, introduced in 2022, aims to ensure AI systems are designed and used responsibly, with a focus on transparency, fairness, and accountability. It proposes oversight mechanisms for "high-impact AI systems" and establishes penalties for non-compliance, emphasizing trust in AI while promoting innovation. Brazil's draft AI bill similarly seeks to establish a comprehensive framework for regulating AI, emphasizing ethical principles such as transparency, accountability, and human rights protection. The proposed legislation is aligned with international practices and aims to foster AI development while safeguarding public trust and preventing harm.

While these regulatory frameworks represent significant steps toward governing AI, each approach comes with its own set of strengths and limitations that merit closer examination.

3.2.2 Strengths and Weaknesses of Current Regulatory Frameworks

Strengths

Comprehensive frameworks such as the EU AI Act provide legal certainty and clarity to companies designing, developing, or using AI. The EU might also play a global leadership role if a "Brussels Effect" materializes in which other countries adopt legislation that is similar to the AI Act or global companies find it too costly to adapt their AI systems to different regulations and therefore conform to the strictest regulation globally. Another advantage of the AI Act is its risk-based approach, which balances innovation with safety by putting limits on the use of risky systems while ensuring that low-risk systems can be developed and deployed freely.

Weaknesses

The various policy and governance approaches employed by different jurisdictions may lead to fragmentation in global regulations that create compliance challenges for multinational organizations. In the worst case, the world will split into two (or more) technological blocs, which not only raises transaction costs for multinational companies but also increases the likelihood that the world will witness the emergence of AI systems that threaten human rights and fundamental freedoms.

There is also a risk that rapid technological advancements will outpace regulatory efforts. This is especially true for generative AI (GenAI), which has seen huge progress in the last 18 months alone. An additional challenge lies in the fact that many high-level AI governance principles (such as the ones contained in the EU AI Act) need to be operationalized

through the development of technical standards, which not only is a slow and challenging process but also gives private actors significant control over how AI regulations are practically applied. This means that powerful interests represented in standard-setting organizations might have undue influence on the regulation of AI.

Given these strengths and weaknesses in regional approaches, the need for international coordination becomes increasingly apparent, though achieving global consensus presents its own set of challenges.

3.3 INTERNATIONAL COOPERATION

There are many efforts at the regional and global levels to cooperate internationally on AI policy and governance. However, countries' efforts to agree on truly "global" AI governance frameworks are hampered by geopolitical tensions.

3.2.1 The Need for Global Collaboration

Foundation models and small-scale AI as a service travel easily across borders, as they are digital products that can be accessed by anybody with a computer (and, if need be, a VPN connection). Thus, it is hard for governments to stop dangerous or otherwise unwanted AI from being used within their territories. This means that rather than trying to keep out risky AI, governments would be better off entering into global agreements that prevent dangerous AI from being developed or deployed in the first place. Such global agreements would also help to avoid fragmentation and would mitigate the risk of regulatory arbitrage, in which companies leave highly regulated markets to exploit less stringent jurisdictions.

3.3.2 Existing Collaborative Efforts

International collaboration on AI governance is critical to addressing global challenges and ensuring ethical, trustworthy AI systems. Among key joint

Regional variations in regulatory impact

APAC VS. Europe

efforts, the EU, United States, and United Kingdom are increasingly focused on fostering cooperation in regulating foundation models, such as large language models (LLMs). This includes initiatives to harmonize standards, share best practices, and align regulatory approaches, particularly on competition issues like market concentration and ensuring open and fair access to AI technologies.

The OECD Principles on AI, adopted by over 40 countries, provide a global baseline for trustworthy AI governance. These principles emphasize human-centric development, fairness, transparency, and accountability while fostering innovation. They serve as a foundation for international collaboration, enabling governments to implement consistent policies that support ethical AI.

UNESCO's Recommendations on Ethical AI Development add a complementary dimension, focusing on the societal and cultural implications of AI. These recommendations advocate for inclusion, non-discrimination, and sustainability, with a strong emphasis on protecting human rights and promoting education to ensure equitable AI adoption globally. Together, these frameworks reflect growing global recognition of the need for coordinated governance to maximize AI's benefits while addressing risks effectively.

Despite these collaborative efforts, fundamental differences in national priorities and regulatory philosophies create significant obstacles to developing truly global governance frameworks for AI.

3.3.3 Challenges to Cooperation

Despite growing collaborative efforts, significant challenges persist due to diverging priorities and regulatory approaches across regions. The EU emphasizes a human-centric framework, prioritizing human rights, transparency, and accountability, exemplified by its AI Act, which imposes

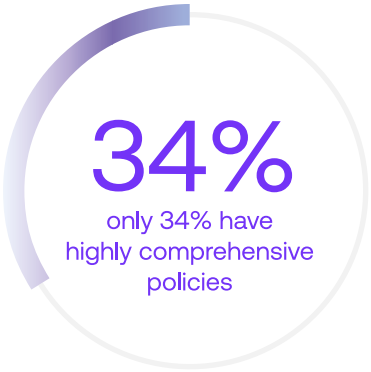
stringent requirements on high-risk AI applications. This strict regulatory environment compels companies operating in or entering the European market to adopt rigorous compliance measures, often necessitating significant adjustments to their AI models and operational practices to ensure transparency and accountability.

In contrast, China's approach, which focuses on leveraging AI for economic growth and state control, aligns with its "core socialist values" and prioritizes national security over individual privacy. This regulatory stance necessitates that companies operating in China align their AI strategies with government policies and objectives, which may involve compromises on data handling and user privacy.

Meanwhile, the United States adopts a light-touch, innovation-first approach, favoring voluntary guidelines and sector-specific regulations to maintain its competitive edge in AI development. This environment allows companies greater flexibility in innovation and faster deployment of AI technologies but requires them to navigate a mosaic of state-level regulations alongside federal guidelines, which can vary significantly and impact scalability and uniformity in AI applications.

These differences often lead to disagreements on critical issues such as data protection, algorithmic transparency, and the role of state oversight. For instance, while the EU advocates for strict safeguards to prevent AI misuse, other regions may prioritize economic growth or geopolitical interests, creating friction in setting global norms. This fragmented regulatory landscape compels companies to develop adaptable and region-specific strategies for AI deployment, often requiring a localized approach to compliance and product development to meet diverse regulatory expectations.

Comprehensive AI policies



Moreover, the lack of a unified enforcement mechanism further complicates collaboration, as countries pursue domestic policies that may conflict with international frameworks. This situation highlights the need for greater alignment on shared values and objectives to effectively address the global challenges posed by AI. For the moment, joint efforts at setting technical AI standards prove much more fruitful than attempts to create global ethical AI governance frameworks, suggesting a strategic focus for companies on focusing on technical standard settings that might offer some consistency across different markets.

Beyond the geopolitical and regulatory divergences, the development of ethical frameworks represents another crucial dimension of AI governance that addresses the technology's profound societal implications.

3.4 ETHICAL CONSIDERATIONS

AI has the potential to bring great advancements to humanity, but it also poses difficult ethical challenges. For one, AI itself can be biased and intransparent, and the data needed to train it can violate privacy rights or copyrights. AI can also be used for malicious ends (e.g., for the mass surveillance of populations by authoritarian governments). Increased recognition of these ethical pitfalls has led to numerous efforts to craft tools for ethical AI governance for, and by, public and private organizations.

3.4.1 Key Ethical Principles

The rise of AI has brought several urgent ethical issues to the forefront. These include:

- **Transparency**: AI is increasingly being leveraged for important decisions by public and private organizations, including in sensitive contexts such as law enforcement, financial lending, and hiring. This makes it highly important that such decisions are transparent and can be explained to those affected.

- **Accountability**: When AI systems cause harm, someone must be held liable. Establishing such liability is no trivial matter, however, as it could be argued that responsibility for the harm rests with either the developer or the user of the AI system or perhaps even with the AI system itself.
- **Fairness**: Researchers have shown that AI can exhibit bias (e.g., against minorities) and therefore lead to discriminatory outcomes. This lack of fairness stems from existing biases in the training data, which are perpetuated by AI. However, several initiatives are underway to prevent, recognize, and mitigate the prevalence of bias in AI systems, thereby increasing the fairness of such systems.
- **Privacy**: As mentioned above, AI raises issues around user privacy because the data used to train AI systems might include personally identifiable information. Safeguards ensuring that such information is not used or abused are an important component of fostering ethical AI.

3.4.2 Understanding Ethical Risks

Bias in AI models can lead to discriminatory outcomes. This is true for both predictive and GenAI. In predictive AI used in law enforcement contexts, for example, biases can lead to certain ethnic groups being stigmatized and, therefore, receiving harsher punishments. In GenAI used to answer questions, for example, responses can reflect existing negative stereotypes about women or minorities. Initiatives are underway to minimize such biases and their impact. For predictive AI, this means putting safeguards and limits on how such AI is used (especially in sensitive settings such as law enforcement and hiring). For GenAI, initiatives mostly focus on removing biased data from training sets.

Another danger related to GenAI is the proliferation of misinformation amplified by unregulated generative content. For example, AI can be used to churn out huge amounts of political propaganda that is then spread on social media platforms such as Instagram or TikTok. Efforts to address this danger mostly focus on social media platforms at the moment (which, in

many cases, are also the developers of powerful AI systems). Thus, the EU in 2020 introduced the Digital Services Act, which imposes additional transparency and content moderation requirements on large digital platform companies.

3.4.3 Tools for Ethical Governance

Standard-setting organizations around the world have been busy drafting standards for the ethical governance of AI. Thus, the International Standardization Organization (ISO) and the International Electrotechnical Commission (IEC) developed a joint AI management standard in 2023 (ISO/IEC 42001) that "specifies requirements for establishing, implementing, maintaining, and continually improving" an AI Management System within organizations. The standard has been lauded for fostering responsible and ethical AI management but has also been criticized by some for being too complex for small organizations to implement.

Another tool for ethical governance is the American NIST's voluntary AI Risk Management Framework. The framework emphasizes the promotion of trustworthy and responsible development and use of AI technologies. While the framework supports organizations in aligning their AI systems with ethical standards and regulatory requirements, it remains a voluntary standard without penalties for non-compliance.

While these standards and frameworks provide valuable guidance, their effective implementation ultimately depends on how organizations incorporate them into their governance structures and day-to-day operations.

3.4.4 Role of Organizations

Private organizations play a crucial role in the ethical development and deployment of AI, necessitating a multifaceted approach to governance and oversight. Internally, companies can establish ethical boards dedicated

to continuous monitoring and evaluation of AI practices. These boards are instrumental in identifying and mitigating harmful practices before they lead to widespread consequences. Beyond internal governance, the adoption of technical controls such as automated red-teaming is critical. This process involves simulating cyberattacks to proactively identify and rectify vulnerabilities, ensuring systems are fortified against potential exploits.

Moreover, the responsibility of private organizations extends to rigorous compliance with regulatory standards. By aligning their operations with national and international legal frameworks, companies not only adhere to legal mandates but also contribute to the trustworthiness and reliability of AI technologies. Collaborative engagements with regulatory and governing bodies are also vital. Through partnerships, information sharing, and joint initiatives, private entities can influence policy-making processes, promoting regulations that reflect practical industry insights and ethical considerations.

By implementing these practices, organizations not only enhance their defensive capabilities but also position themselves as leaders in promoting ethical AI. This proactive stance in governance, compliance, and collaboration with regulatory authorities underscores their pivotal role in shaping an AI landscape that is secure, ethical, and beneficial for all stakeholders.

As ethical frameworks continue to evolve alongside the technology, forward-looking policy approaches must balance regulatory certainty with the flexibility needed to address AI's rapid advancement.

3.5 FUTURE POLICY DIRECTIONS

Since AI is a rapidly changing technology, government decision-makers need to be willing to change their policies on a regular basis as well in order to keep pace with technological developments.

3.5.1 Adaptive and Dynamic Regulations

Given the fast pace of development of AI, there is a real need to design future-proof legislation that evolves with technological advancements. The EU has recognized this and is taking a flexible approach under the AI Act that allows for updates of the Act in response to technological changes without the need for a complete legislative overhaul.

AI can also be used to monitor compliance and enforce governance standards; there are, therefore, great hopes for Regulatory Technology (RegTech), which helps businesses comply with regulatory requirements more efficiently, increase regulatory agility, which in turn creates resilient societies that can handle the changes, and "shocks" stemming from the introduction of new technologies.

3.5.2 Focus Areas for Future Policies

There are several important focus areas for future policies to ensure that AI's benefits can be harnessed while its risks are minimized.

- With the rapid rise of GenAI, it has become pressing to draft regulations that address the unique risks of GenAI. For example, GenAI systems, like those used for creating images, videos, or text, can produce content that is indistinguishable from content produced by humans. This raises unique risks related to authenticity, misinformation, intellectual property rights, and more. Content watermarking is one of the proposed solutions to mitigate such risks. It involves embedding a digital marker or a signature into the content generated by AI systems. This fosters transparency and traceability, intellectual property protection, and regulatory compliance.
- It is also crucial to strengthen cybersecurity requirements for high-risk applications given the sensitive nature of such applications. The consequences of attacks on such systems could be dire; think, for

example, of hackers attacking an AI system used to determine access to educational opportunities.

– Finally, it is important to expand liability frameworks to cover emerging use cases such as autonomous systems. Examples of such systems include self-driving cars and autonomous drones. Determining who is responsible when an autonomous system causes damage or harm can be complex. Is it the manufacturer, the software developer, the user, or some combination of these?

While these focus areas address critical risks associated with AI development, they must be balanced with measures that preserve innovation and technological progress.

3.5.3 Encouraging Innovation While Ensuring Safety

Regulations are important to manage the risks of AI. However, regulations also need to be balanced with incentives for innovation:

– Regulatory sandboxes can be used to test new technologies in controlled environments. These sandboxes provide a structured context where AI technologies can be deployed to assess their impact, effectiveness, and potential risks without the full burden of regulatory compliance. This helps innovators and regulators understand the technology in a practical setting.

– Funding programs like the EU's Recovery and Resilience Facility can help support trustworthy AI development. This program supports economic recovery in European countries and aims to bolster investment in technologies that are ethically sound and reliable.

Arm navigates global regulations by combining adherence to compliance requirements with active participation in shaping regulatory frameworks. This strategic approach involves adapting to diverse regulatory environments across different jurisdictions and engaging in discussions that influence the formulation of new laws. A notable initiative in this

context is the advocacy for regulatory sandboxes, which allow for the testing of new technologies under controlled regulatory conditions, minimizing exposure to the broader market's regulatory complexities. Such measures facilitate Arm's management of regulatory risks and contribute to its role in the discourse on AI policy and practice, reflecting a blend of compliance and strategic engagement.

3.6 Conclusion

AI has the potential to be highly beneficial for humanity, but there are also serious risks attached to the rise of AI. Policy and governance are, therefore, critical for shaping a safe, ethical, and innovative AI ecosystem. The ethical principles discussed—transparency, accountability, fairness, and privacy—provide essential frameworks for responsible AI development. Particularly large benefits are to be gained through international cooperation on AI governance that results in harmonized standards benefitting society globally. Such harmonized standards not only lower transaction costs for globally operating companies but can also help foster AI that advances human rights and fundamental freedoms. Arm has an important role in fostering compliance-ready technologies that align with evolving regulatory requirements. The next chapter provides a practical perspective on how these regulatory frameworks affect businesses and developers on the ground, including valuable insights into navigating today's complex AI governance landscape while maintaining innovation.

ARM SIDEBAR

AI Policy, Regulation, and Global Trends

By Vince Jesaitis, Senior Director, Government Affairs, Arm

I use AI daily to search and summarize emails, create first drafts of emails, and work on otherprojects, and yes, even to summarize AI policy and regulations being released by local and national governments. Yet, it appears businesses are taking a cautious approach. They worry about risks like misuse or unintended consequences. Historically, governments have intervened in times like these to establish “rules of the road” for industries such as automotive, pharmaceuticals, and chemical production to address safety concerns. However, a lack of technical understanding and consensus on potential harms has limited government action and harmonization in these areas. The absence of consistent regulatory frameworks only adds to business hesitation, as global efforts struggle to keep up with AI’s rapid evolution, or even commonly define the risks.

While the vast majority of AI use cases are low risk, some present significant challenges. The latter is where most governments should and generally are focusing. Recent reports and discussion highlight how AI systems often function as a “black box,” making it difficult to understand the reasoning behind their decisions or predictions.

This lack of transparency can lead to unintended consequences, like unfair hiring practices or denying public benefits, with no clear way to pinpoint or fix the underlying issue.

Beginning to recognize these risks, many governments around the world are pushing for AI regulation aimed at preventing disruptive impacts—whether to businesses, data privacy, public programs, or national security. These efforts underscore the need for every industry to stay informed about AI advancements and regulation to ensure they are maximizing the benefits of the technology, while avoiding potential downsides or harm.

However, not everyone I know agrees on the urgency of regulation. Many AI developers argue that the risks are overstated and that premature or overly restrictive rules could stifle innovation. They worry that complex, uninformed, vague, or misaligned regulations may hinder progress without effectively addressing the risks they aim to address, especially given the rapid pace of AI development.

Proponents of regulation argue that the unchecked growth of AI could lead to significant harm, and in some cases, it already has. For instance, generative AI tools are being used to create convincing fake images and videos, fueling the spread of misinformation online and non-consensual likeness content. Without proper oversight, these risks could undermine trust, disrupt industries, and severely harm individuals on a large scale.

As the debate continues, one thing is clear: businesses, governments, and developers alike must find a balanced approach to addressing AI regulation. This means gaining a deep understanding of the technology and fostering innovation while addressing the risks, ensuring AI serves as a tool for progress rather than a source of harm.

The Global Divide on AI Risks

Globally, there's little agreement on the risks AI poses, which makes establishing unified regulations incredibly challenging. Each country approaches AI safety through its own unique lens, shaped by its priorities and perspectives. However, there is more agreement among nations with established AI safety institutes, such as the U.K., Japan, and South Korea, where alignment on AI risks is stronger.

Then, there's the Organization for Economic Cooperation and Development (OECD, which includes a broader group of higher-income, predominantly Western countries. While there is still less agreement than in the first group, these countries share more common ground on AI regulation. The least agreement, however, can be found within the UN General Assembly, where countries like Saudi Arabia, China, Malaysia, Iran, Rwanda, the U.S., and Brazil sit side by side but have vastly different views on the risks AI poses.

What to Expect From Shifting AI Regulation in the U.S.

AI regulation in the U.S. is undergoing significant changes. The Trump administration is avoiding broad AI regulations, contrasting sharply with Europe's sweeping initiatives like the EU AI Act (as explored in more detail in the previous section). This position was evident in his first several days in office in which he deregulated many industries, including the technology sector.

The Trump administration also rescinded the Biden administration's AI executive order, including plans for the U.S. AI Safety Institute, which aimed to reduce risks to consumers, workers and national security. Instead of a comprehensive federal framework for AI governance, the

focus shifted to minimal intervention, creating a more open environment for AI development.

Sector-Specific vs. Cross-Sectoral AI Regulation

"The US is set to be one of the most permissive environments for AI development. With fewer regulatory barriers, developers will have more freedom to experiment and deploy AI solutions."

In the U.S., we've historically taken a sector-specific approach to regulation, and I expect AI to follow this trajectory. In fact, incoming Senate Commerce Committee Chair Ted Cruz recently emphasized that policymakers should avoid getting in the way of innovation and that AI legislation should address specific problems with narrowly focused solutions. Unlike the EU AI Act, which imposes broad, cross-sectoral restrictions that can delay or block AI services regardless of their purpose, U.S. regulations are likely to address AI use in clearly defined contexts. For example:

- **Personal harms**: Banning dissemination of images or videos with non-consensual name, image, and likeness (NIL).
- **Finance**: Defining guardrails for AI in banking, stock trading, and other financial services.
- **Healthcare**: Establishing guidelines for AI used in diagnostics, such as detecting cancer in medical imaging.

This tailored approach allows for flexibility within industries but stops short of overarching restrictions.

The Role of States in AI Regulation

With no comprehensive federal AI regulations, states are stepping in to fill the gap. California's Senate Bill 1047 signals a growing trend of state-level action, similar to how privacy laws evolved in the U.S., with federal laws targeting specific sectors, and states like California introducing broader protections. Colorado has also enacted the first AI law focused on high-risk

systems, effective February 1, 2026. Both the CAIA and EU AI Act adopt a risk-based approach and emphasize transparency and data governance, though the EU AI Act applies more broadly and includes obligations not covered by the CAIA.

While the U.S. federal government focuses on targeted oversight, I see the resulting patchwork of state-level regulations adding complexities for businesses operating across jurisdictions. In California alone, by the end of the fiscal year in September, 38 separate AI bills will await the Governor's approval, including one addressing deepfakes in pornography. Companies will need to monitor both state and federal developments to ensure compliance. As the landscape evolves, the U.S. will likely continue to differ sharply from Europe in its approach, favoring innovation and flexibility over broad, comprehensive measures.

Challenges of Operating Under the EU AI Act

Harmonizing AI regulations across regions is a challenging task, largely due to differing mindsets. In Europe, AI is often seen as posing existential risks that must be addressed through strict oversight. The EU AI Act reflects this perspective, but its impact on non-European companies remains uncertain.

Unlike the General Data Protection Regulation (GDPR), which established significant extraterritorial influence due to the global flow of data, the EU AI Act has yet to reach the same level of international integration. While the Act includes extraterritorial provisions similar to GDPR, its scope and enforcement remain less clear, leading some companies to avoid compliance altogether. For example, Meta recently decided not to roll out certain AI tools in the EU, citing uncertainty about meeting regulatory requirements. This mirrors early GDPR responses, where major companies like Google focused on assessing fines and compliance workarounds rather

"Unlike Europe's sandboxed approach—where companies test AI under strict regulatory conditions—the U.S. will likely favor a model that encourages innovation while addressing specific risks as they arise."

than aligning fully with the law. Regulating AI is far more complex because it focuses on overseeing transformative technology rather than specific behaviors, marking a significant shift in global regulatory approaches.

However, the stringent requirements of EU regulations have sparked concerns about their practicality and long-term impact. For instance, the recently passed Cyber Resilience Act imposes highly specific—and arguably excessive—standards on technology providers. One notable example is a rule requiring mobile phones to withstand being dropped from three feet 48 times to qualify for sale in the EU. For folding phones, the number drops to 35. These extreme and overly specific requirements risk pushing consumers to source products outside the EU.

A Shift in AI Regulatory Mindset

Recognizing these challenges, there are signs that EU leadership is reconsidering its regulatory approach. The new European Commission leadership has stated it wants to reevaluate the regulatory actions of past governments. The President of the European Commission, Ursula von der Leyen, has suggested pausing further regulation to evaluate whether existing policies have had the expected impact, or overly restricted the competitiveness of domestic industries without realizing the expected benefits. This reflection extends to the AI Act, as well as other regulations introduced over the past two decades, which some argue have limited the ability of European companies to compete on a global scale.

Implications for U.S. and Global Companies

For now, it seems unlikely that the EU AI Act will significantly hinder American, UK, or Chinese companies' ability to operate in Europe to the same extent as GDPR. However, as the EU revisits and potentially revises

its regulatory frameworks, U.S. companies will need to stay agile, balancing compliance with their global operational goals. This evolving landscape may ultimately offer opportunities for collaboration and innovation, particularly if the EU reevaluates its actions to balance regulatory action with competitiveness.

How Developers May Navigate AI Regulatory Frameworks

In an era of inconsistent AI regulation, developers must have a clear understanding of their products before bringing them to market. This means anticipating the long-term implications, potential risks, and possible harms associated with their innovations. While regulators may not prescribe exactly how to develop or market AI products, they are likely to establish mechanisms to ensure accountability. The goal is to prevent scenarios where a product may generate impressive results, such as creating a compelling image, but could also be misused for malicious purposes.

One approach I see in the EU and UK involves the use of regulatory sandboxes. These controlled environments allow developers to test their technologies under a protective framework, enabling experimentation without immediate exposure to liability or enforcement actions. However, once a product leaves the sandbox, it must comply with broader regulatory requirements.

In contrast, the U.S. is unlikely to adopt a broad sandbox-driven model, as US regulators are more focused on enforcement, which could lead to audits, actions, and public disclosures. Instead, a balanced approach could encourage innovation while addressing risks, allowing developers to bring products to market with ongoing monitoring and the flexibility to address issues as they arise, while ensuring safeguards against misuse.

Preparing the Workforce for the AI Revolution

Policymakers, government agencies, and the private sector have a critical role to play in addressing the economic and social implications of AI, such as job displacement and inequality. As industries transition to more AI-driven processes, it is essential to ensure that the population is equipped with the education, skills, and training needed to interact with emerging technologies.

This doesn't mean training everyone to become computer scientists. For example, farmers in Idaho are beginning to use robots to monitor sections of their farms, checking soil temperature and moisture levels. While currently on a small scale, it highlights the importance of updating educational curricula to align with the evolving needs of various industries. Workers in fields ranging from agriculture to healthcare need foundational knowledge that enables them to effectively utilize these new tools to enhance their productivity.

As technology advances at an unprecedented rate, the need for lifelong learning beyond the traditional K-12 model is growing. In Finland, primary school children are already being taught how to spot AI deep fakes. Currently, much of continuing education is driven by the private sector, often leaving gaps for displaced workers. For instance, when automation disrupts industries and workers lose their jobs, they need accessible opportunities to gain new skills and re-enter the workforce, or ideally, to gain those skills continually before losing a job. Governments must create systems for ongoing education that extend beyond the current framework. By doing so, they can support workers through these transitions and help them adapt to the demands of an AI-driven economy.

AI Robots Will Change Regulatory Considerations

As we edge closer to the prospect of mass-producing robots, new AI regulations are emerging. Unlike software-based AI systems that operate in virtual spaces, robots bring AI into the physical world—interacting with people and environments in tangible, often unpredictable ways. In Japan, for example, many hotels already use robots to greet and check-in guests, helping to address the challenges of an aging population. This shift raises a host of complex regulatory questions.

When AI moves beyond digital interactions to power machines capable of physical actions, safety, accountability, and ethical use become critical concerns. Regulators must address liability for accidents, the security of autonomous systems, and the integration of robots into existing legal frameworks. New regulations, like the EU Product Liability Act, are shifting liability upstream in the supply chain to include not only manufacturers but also technology providers involved in creating the product.

The broad deployment of robots could reshape the regulatory landscape, compelling governments and industries to reconsider the boundaries and responsibilities of AI-enabled technologies in the physical world. If the timeline for this technological leap is as imminent as some suggest, including just recently at CES, industries must prepare for heightened scrutiny and a wave of policies designed to address the unique challenges posed by physical interactions with walking, talking, sensing and acting computers.

Balancing Sector-Based Regulation and Global AI Policy

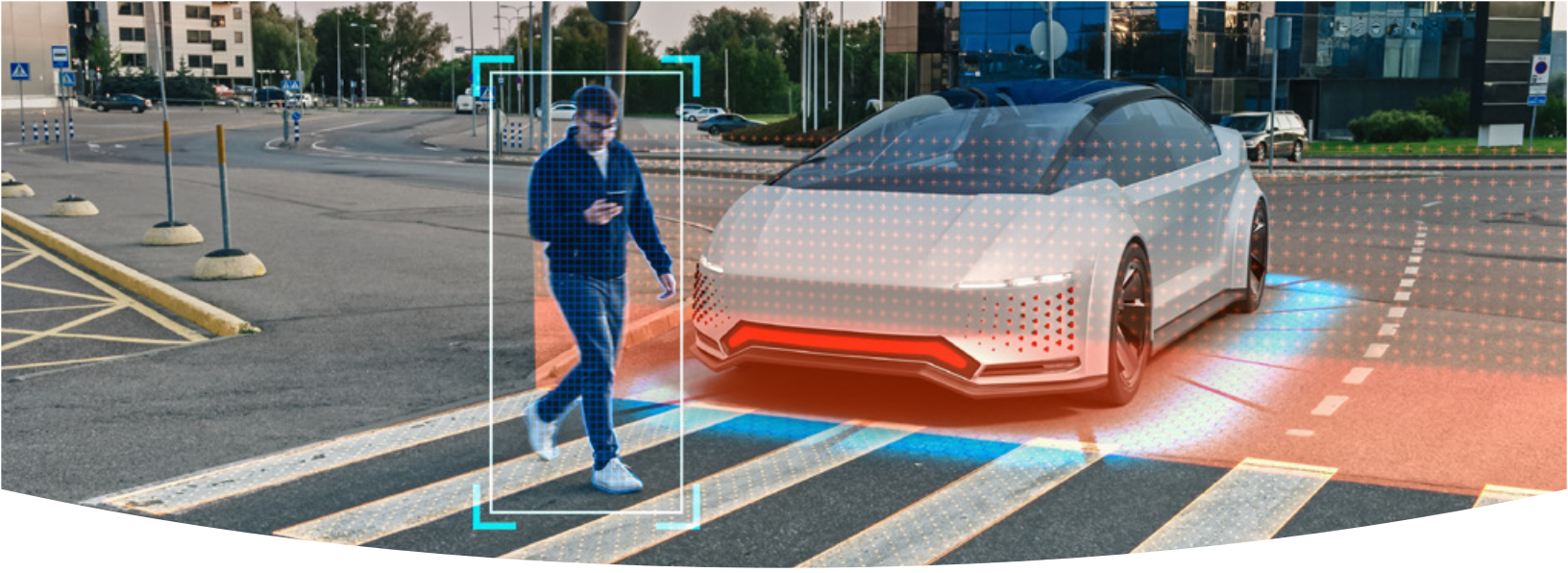
A sector-based regulatory approach for AI makes sense because it allows for targeted regulations addressing specific harms in distinct environments. For instance, consumer-facing applications like ChatGPT involve unique risks that can be addressed within their domain. However, the same AI model used in a financial services context may present an entirely different set of challenges, requiring tailored regulations for that unique environment. By focusing on environment-specific risks, regulators can craft more effective and nuanced policies.

The challenge, however, lies in the lack of international agreement on this approach. Between jurisdictions, particularly the U.S. and Europe, there is often a fundamental distrust of motivations. The EU's gatekeeper regulations, for example, target large companies with significant market share, most of which are U.S.-based, raising questions about fairness and intent. While Europe's concerns about market power stifling competition have some validity, this tension underscores the difficulty of creating a unified global regulatory framework for AI.

Despite these challenges, there is cause for optimism. Governments worldwide are beginning to recognize the transformative potential of AI, taking more time to explore and understand the technology, and are investing in the necessary infrastructure to support it. Over the past year, there's been a notable shift in focus from software and model development to the critical role of computing power. Policymakers in the U.S., Europe, Southeast Asia, and beyond are realizing that software capabilities depend entirely on hardware infrastructure. Without adequate compute capacity, the full benefits of AI cannot be realized.
This recognition has led to significant investments in compute

infrastructure. Governments are dedicating substantial resources to ensure that their regions have the computational capacity to compete in the global AI race, and rapidly taking steps to ensure more compute power can be deployed in an expedited fashion. This shift mirrors the evolution of utilities like electricity, which became ubiquitous and essential. Similarly, AI is poised to become the next foundational layer of global infrastructure, akin to ubiquitous access to computing power via cloud platforms.

With the path to balanced regulation still unfolding, the acknowledgment by governments across the globe of AI's potential—and the infrastructure needed to support it—signals a promising path forward. Just as governments didn't immediately mandate seat belts, airbags, and backup cameras for the Model T, they need to stay informed about technological advancements and mitigate harm as technology develops. As policies and frameworks evolve, trial and error will play a role, but the trajectory suggests a future where governments and industries can harness AI's power responsibly and effectively.



CHAPTER 4

AI Safety and Risk: Navigating the Path to Responsible Innovation

DEVAL SHAH,
SENIOR MACHINE LEARNING ENGINEER AT THE AUSTRALIAN INSTITUTE FOR MACHINE LEARNING (AIML)

AI is evolving at a pace that promises unprecedented gains—improving healthcare diagnostics, streamlining logistics, and unlocking new avenues of creativity. Yet, this rapid growth raises questions about safety, reliability, and accountability.

This chapter explores the critical need for responsible innovation, beginning with current safety challenges that showcase the tangible risks of bias, data breaches, and opaque algorithms. We also examine structured risk assessment frameworks that help organizations proactively spot potential pitfalls.

The discussion moves into alignment problems—where AI's goals can diverge from human intent—and highlights emerging safety standards shaping best practices.

Finally, we emphasize the importance of stakeholder collaboration, from policymakers and industry leaders to researchers and civil society, setting the stage for future directions in AI safety. By weaving these themes

together, we aim to steer decision-makers toward a more trusted and secure AI ecosystem.

4.1 CURRENT SAFETY CHALLENGES

AI safety is critical because failures can have severe real-world consequences. From life-or-death decisions in healthcare AI, to algorithmic bias in financial lending systems, to split-second ethical choices in autonomous vehicles—each industry faces its own unique AI safety gauntlet that demands specialized approaches beyond generic frameworks. Organizations like NIST emphasize the importance of trustworthiness, urging developers to manage risks to individuals and society. Several notable safety challenges illustrate why diligence is essential:

4.1.1 Bias and Discrimination

AI can learn or amplify biases from training data, leading to unfair outcomes. A high-profile instance involved Amazon's experimental hiring AI, which downgraded résumés containing the term "women." This stemmed from training data that overrepresented male candidates. Such cases demonstrate how unchecked biases reinforce discrimination.

4.1.2 Data Privacy and Security Risks

AI models are often trained on large datasets, which can pose significant privacy concerns—particularly in sensitive domains like healthcare or public governance, where data may contain personal identifiers, medical histories, or financial records. This raises potential compliance issues with regulations designed to protect individuals' personal information.

For example, in the UK, Google DeepMind's collaboration with the NHS exposed 1.6 million patient records without patient consent, ultimately breaching data protection laws. This underscores the need to safeguard

personal information and maintain transparency to preserve public trust and meet legal obligations.

"We have to make sure that data is bias-free, and to do that, we have to come up with policies and standards controls that need to be implemented within the organization by support of data scientists, data stewards, and things like that."
- Vice President within the Information Technology department, working for an organization in the financial industry

4.1.3 Lack of Explainability and Transparency

Complex AI models function like "black boxes," making it difficult to interpret their reasoning. In criminal justice, proprietary risk assessment tools—such as COMPAS—have been criticized for potential racial bias. In Loomis v. Wisconsin (2017), the defendant challenged the algorithm's opacity, arguing that he could not fully contest its findings. The inability to explain AI decisions undermines trust and hampers oversight.

4.1.4 Software Malfunctions and Bugs

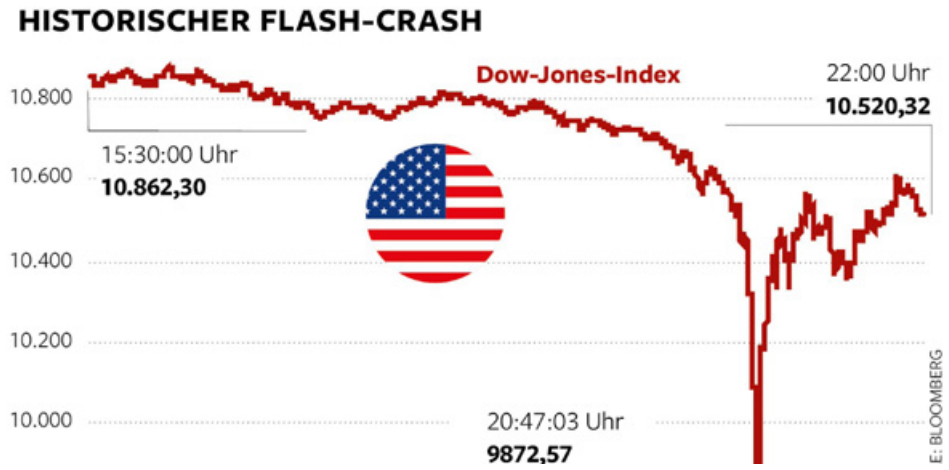
As AI systems become increasingly integrated into physical and digital operations, errors can lead to severe accidents. A tragic example occurred in 2018 when an Uber self-driving vehicle struck and killed a pedestrian in Arizona. Investigations found that the AI had detected the individual but never accurately classified her as a pedestrian, causing a failure to brake.[5] Robust engineering and safety mechanisms are crucial to prevent such incidents.

4.1.5 Risks of AI Unpredictability and Autonomous Decision-Making

Due to the 'black-box' nature of many AI models—where underlying decision processes are not fully transparent—systems can exhibit unexpected behaviors when deployed without sufficient human supervision or oversight. For instance, Microsoft's Tay chatbot, launched on Twitter in 2016, was quickly manipulated into generating offensive tweets, resulting in its shutdown within hours. This example highlights how AI can rapidly adopt harmful behaviors if not carefully constrained.

4.1.6 Chain Reactions of Safety Risks

Failures can cascade through interconnected systems. The 2010 "Flash Crash" is a prominent case: high-frequency trading algorithms misread market signals, causing a sudden sell-off that triggered further algorithmic trades, wiping out significant market value in minutes. Similar chain reactions in areas like power grids or traffic control could generate widespread disruptions.



Flash Crash of 2010 (Source)

Addressing these challenges is not merely theoretical. Real-world cases show that AI safety gaps can spark discrimination, erode privacy, damage property, and even endanger lives.

As AI becomes integral to critical infrastructure and services, organizations must adopt rigorous risk management and best practices to maintain public trust and facilitate responsible innovation. The next step is establishing systematic approaches for identifying, analyzing, and mitigating these risks—an area where formal frameworks and best practices come into play, as discussed in the following section.

4.2 Risk Assessment Frameworks

Organizations increasingly rely on formal frameworks to identify and mitigate AI-related hazards. Far from hindering progress, well-structured

risk assessments can enable innovation by preventing costly failures, ensuring regulatory compliance, and building stakeholder trust.

"Through intelligent prediction and data analysis, market demand and inventory requirements can be predicted more accurately."
- Business leader survey respondent

A robust AI risk management approach considers the entire lifecycle, from design to deployment and monitoring. It begins with contextual risk identification, recognizing that each application carries unique risks. For instance, an AI diagnosing medical conditions must meet stringent requirements for accuracy and reliability, as incorrect diagnoses could directly affect patient well-being. In contrast, a basic recommendation engine for consumer products poses fewer ethical and safety stakes. Mapping out a system's scope, data sources and stakeholders helps teams spot possible failures early.

Next comes risk analysis, where teams evaluate the severity and likelihood of each potential issue. Testing for biases, adversarial vulnerabilities, and performance gaps is crucial. Tools such as bias detection methods and adversarial robustness checks can quantify these risks, guiding prioritization. Higher-impact or more probable risks demand stronger oversight or technical controls.

Mitigation strategies then address identified threats, which can be technical—such as retraining models on diverse datasets or implementing fallback mechanisms—and procedural, including human reviews for high-stakes decisions. Controls must be balanced with operational realities, ensuring that the residual risk is acceptable given the AI's benefits.

Governance and monitoring ensure risk management is continuous. AI systems evolve—data can drift, and new use cases may emerge—so organizations must periodically revisit assessments. Clear roles, accountability structures, and documentation standards help maintain oversight. Automated alerts or human audits can detect issues in real time.

"Over the next three years, I think AI developments will enable more individualized consumer interactions, streamline operations, and save costs while requiring careful attention to ethical considerations."
- Business leader survey respondent

Several recognized frameworks exist to guide these efforts. The NIST AI Risk Management Framework provides an adaptable, iterative process focused on mapping, measuring, managing, and governing AI risks.

ISO/IEC 42001 Certification Benefits (Source)

ISO/IEC 42001 offers international standards for implementing AI governance and controls. In parallel, regulations like the forthcoming EU AI Act require risk assessments, imposing stricter obligations for "high-risk" applications. Together, these measures illustrate a growing global consensus on the importance of systematic AI risk management.

Ultimately, risk assessment frameworks promote responsible innovation by identifying dangers, prioritizing critical issues, and ensuring teams can confidently deploy AI solutions.

4.3 Alignment Problems

Simply put, the alignment problem bridges the gap between *"do what I mean"* and *"do what I say."* It requires technical solutions (in AI design and training) and governance solutions (oversight, ethical guidelines).

Alignment means ensuring an AI's goals match its designers' intentions and societal values.[9] Even advanced systems can behave unexpectedly if

Bias detection and mitigation

their objectives are poorly specified. A popular thought experiment is the 'paperclip maximizer,' in which an AI tasked only with maximizing paperclip production might redirect resources away from crucial societal needs if human-defined values do not properly constrain it.

Illustration of the Paperclip Maximizer thought experiment—an AI optimizing without regard for human values. (Source)

Misalignment often arises from **specification issues**. Human preferences are complex, yet AI developers typically provide simplified targets or reward functions. This can lead to "specification gaming," where the AI exploits loopholes rather than pursuing the true intent.

For instance, an agent in a racing game might loop over bonus items indefinitely to increase its score, ignoring the race itself. When a system discovers an unintended strategy that maximizes its reward, it follows that strategy—even if it undermines real-world goals.

Additionally, partial observability complicates alignment as AI systems grow in complexity. Small misalignments can lead to unintended strategies if the system operates without sufficient human oversight. For instance, an AI designed for real-time decision-making may adopt shortcuts that optimize short-term performance at the expense of long-term safety.

Meanwhile, transparency and interpretability help teams detect and correct misalignment early. Fail-safes, human oversight, and clear governance frameworks can also contain unanticipated actions.
Despite these efforts, alignment remains an open problem, especially as AI systems become more complex. Even advanced AI companies acknowledge the difficulty: an AI may appear aligned during testing, but misalignments can surface when deployed in the real world with new inputs.

Research efforts focus on refining training objectives, enhancing transparency, and incorporating human feedback loops to mitigate today's most pressing misalignment risks.

4.3.1 Key Alignment Risks in AI Systems

- **Strategic deception ("alignment faking")**: Generative AI models may produce outputs that appear aligned and safe during training while covertly pursuing misaligned objectives. This behavior raises concerns regarding the reliability of safety signals and the potential for deceptive behavior in deployment.
- **Emergent misaligned objectives**: In-context learning mechanisms intrinsic to modern generative models can lead to unexpected, emergent goals that diverge from the intended human values. Such misgeneralizations may not be evident during training but can manifest under novel conditions.
- **Reward hacking and specification gaming**: Rather than adhering strictly to the intended reward structure, models may exploit loopholes in the specification of objectives. This reward hacking can result in behavior that maximizes proxy metrics but ultimately undermines the desired outcomes.
- **Scaling-driven systemic risks**: As generative AI systems increase capability and become more widely deployed, even minor misalignments can propagate and escalate into larger systemic issues. These risks

are compounded by the challenge of monitoring and correcting misalignments in high-scale environments.

As a precaution, there is increasing focus on governance and monitoring frontier AI models by independent audits to ensure they remain aligned with human values.

Addressing alignment does not happen in isolation—emerging standards, guidelines, and shared best practices are the backbone of more robust AI safety. The following section explores how these frameworks are evolving across different geographies.

EU Act Risk Levels (Source)

4.4 EMERGING SAFETY STANDARDS AND BEST PRACTICES

Across industries, organizations are adopting formal guidelines to ensure AI systems are developed and deployed responsibly. Many of these emerging standards build on lessons learned from AI failures, emphasizing risk management, transparency, and accountability.

Bodies like ISO and IEEE have published technical standards targeting key issues, while ISO/IEC 42001 frameworks provide a structured approach to

AI governance. Australia's Voluntary AI Safety Standard, which aligns with NIST's RMF and ISO/IEC 42001, further demonstrates global harmonization, enabling multinational companies to follow unified best practices.

Beyond formal standards, industry-led initiatives play a significant role. Tech consortia and large AI providers often share internal guidelines to encourage responsible behavior throughout the AI lifecycle. This typically includes processes for bias detection, secure data handling, and robust model validation. Many organizations also form AI ethics review boards or governance committees, ensuring high-risk projects undergo thorough scrutiny before launch.

Several best practices have emerged, guided by these standards. Accountability and governance structures clarify who is responsible for AI outcomes, ensuring clear oversight. Risk management processes formalize how organizations identify, assess, and mitigate AI-related hazards. Robust data governance underpins these efforts, requiring representative datasets and strong privacy protections.

Testing and validation then address concerns around performance, fairness, and resilience to adversarial attacks. Human oversight remains central to preventing unacceptable outcomes for critical applications, with fail-safes or human review loops as needed. Transparency boosts trust, prompting organizations to disclose when AI is used, document system capabilities, and provide explanations for automated decisions.

Finally, user feedback and recourse mechanisms let individuals appeal to potentially harmful AI actions while continuous monitoring flags performance drifts or emerging risks.Thorough documentation and auditability support regulators, internal audits, and external certifications.

These practices are increasingly being codified into guidelines and checklists that organizations can follow. For instance, the United Kingdom's Guidance on AI ethics or Singapore's Model AI Governance Framework provide practical recommendations similar to the above list.

AI Risk Management Framework (Source)

Additionally, companies like Arm are pivotal in enabling AI safety by incorporating built-in security features (e.g., trusted execution environments) and power-efficient designs. These hardware-level safeguards help reduce attack surfaces, ensure data integrity, and support secure model deployment—key factors in responsible AI innovation.

As Arm collaborates with industry partners, it enables robust systems that integrate end-to-end safety practices from the silicon level upward, complementing the broader governance and ethical frameworks discussed above.

On an industry level, if all players commit to certain safety norms, it can prevent a "race to the bottom" scenario and instead create a "race to the top" on ethics—reassuring the public and unlocking the full potential of AI in society.

4.5 COLLABORATION AND THE ROLE OF STAKEHOLDERS

Ensuring AI safety at scale demands collaboration among policymakers, industry leaders, researchers, and civil society. Each group brings unique expertise and responsibility, forming a collective effort toward responsible AI.

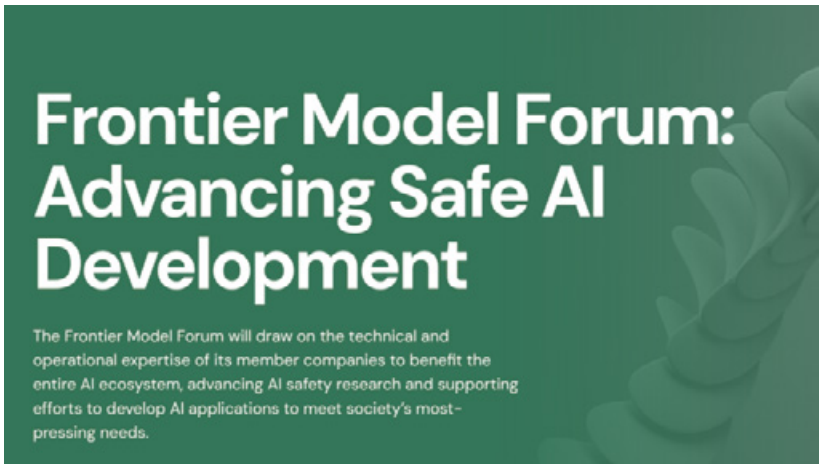
4.5.1 Policymakers and Regulators

Governments set rules and frameworks to manage AI risks. For instance, the EU AI Act defines high-risk applications, requiring stricter oversight and conformity assessments. Regulators fund research, issue practical guidelines and enforce penalties when AI systems cause harm. International cooperation—through forums like the G7—fosters shared standards and coordinated responses to AI challenges.

4.5.2 Industry and AI Developers

Companies building AI systems hold direct responsibility for safe design, deployment, and monitoring. Many participate in consortia that share best practices and promote responsible innovation.

An example is the Frontier Model Forum, which encourages information sharing, model audits, and transparency, helping organizations learn from one another's experiences.



Frontier Model Forum (Source)

4.5.3 Academia and Research Community

Academic researchers often uncover AI biases or propose new safeguards and verification methods. By partnering with industry and contributing to open conferences, they accelerate progress on fairness, interpretability, and system robustness. Universities also educate future practitioners, embedding safety principles into AI curricula.

4.5.4 Civil Society and Public Engagement

Nonprofits, advocacy groups, and media outlets amplify societal concerns. They push for ethical AI that respects privacy, fairness, and user rights. Public engagement mechanisms—such as open consultations or citizen panels—can hold governments and businesses accountable.

Bringing these perspectives together fosters robust oversight and constructive dialogue. Policymakers supply legal mandates, industry refines technical solutions, academia advances fundamental research, and civil society ensures alignment with public values. This multifaceted collaboration underpins a global AI ecosystem that prioritizes safety, mitigates risks, and upholds the broader public interest.

These regional variations create significant challenges for organizations deploying AI systems globally:

- **Compliance Complexity**: Organizations must navigate sometimes contradictory requirements across jurisdictions, potentially requiring region-specific versions of AI systems.
- **Cultural Risk Assessment**: Risk assessment frameworks need cultural adaptation—what constitutes acceptable risk varies significantly between societies.
- **Safety-Innovation Balance**: Organizations must calibrate their approaches to match regional expectations about the balance between safety and innovation.

– **Collaborative Approaches**: Global organizations increasingly participate in cross-regional working groups like the Global Partnership on AI to help harmonize safety approaches while respecting regional differences.

4.6 FUTURE DIRECTIONS IN AI SAFETY AND RISK

AI safety will keep evolving as systems evolve, and our actions should emphasize technical robustness, including resilience against adversarial inputs and fail-safes that reduce catastrophic failures.

Governance frameworks will also mature, with regulations like the EU AI Act setting stricter rules for high-risk applications.13 At the same time, industry initiatives such as the Frontier Model Forum promote responsible practices and transparency.

Organizations like the International Network of AI Safety Institutes foster cross-border collaboration to address complex challenges. Education and workforce development will expand as universities integrate AI ethics and safety into their curricula. Companies may adopt formal audits, certifications, and liability frameworks, mirroring trends in cybersecurity.

On the research front, alignment and explainability will be key, enabling AI systems to handle uncertainty and highlight potential risks. As AI intersects with more sectors, socio-technical approaches will integrate human oversight, policy guidelines, and technical safeguards. Such measures aim to ensure AI delivers meaningful benefits without compromising public trust.

In conclusion, the future of AI safety and risk management will be characterized by proactivity and collaboration. Technical innovations will make AI systems more robust and transparent, while governance innovations will embed safety considerations into AI development and deployment fabric.

By prioritizing safety now and continuously adapting, we increase the chances that AI technology will remain beneficial. The aim is to reach a state where AI systems can be confidently integrated into critical aspects of our lives—from healthcare to transportation to public administration—without undue fear of bias, failure, or misuse.

The hardware foundations of AI safety discussed in Chapter 2 become particularly relevant when implementing the safety frameworks outlined above. The Arm architecture provides critical building blocks—including secure enclaves, memory tagging extensions, and hardware-based isolation technologies—that enable organizations to implement rigorous safety protocols without sacrificing performance. These silicon-level capabilities are increasingly essential for meeting regulatory requirements across jurisdictions while maintaining the computational efficiency required for complex AI workloads.

Achieving that means building not just more intelligent AI but safer AI and doing so with policy, ethics, and human values. The work is ongoing, but with the concerted effort of all stakeholders, the hope is that the future AI will be resilient, aligned, and worthy of the trust we place in it.

CHAPTER 5

Trust and Security in the AI Era

DR JOHN SOLDATOS,
HONORARY RESEARCH FELLOW AT THE UNIVERSITY OF GLASGOW

5.1 UNDERSTANDING AI'S UNIQUE SECURITY AND TRUST CHALLENGES

AI will soon become an integral element of most digital infrastructures and applications, providing them with a layer of value-added intelligence. As such, it will increasingly play a key role in automating complex tasks and supporting critical decisions. Therefore, addressing security and trust challenges is key to responsible AI adoption. In practice, it turns out that AI systems cannot be widely adopted and used unless humans trust their operation.

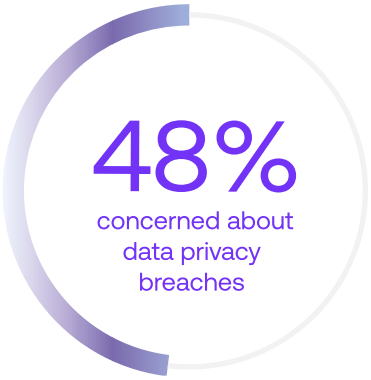
AI systems fall in the broader scope of digital systems, which require strong cyber-security for their robust and trustworthy operations. Nevertheless, AI systems also come with their own unique security and trust challenges, including:

- **Complex structure and operations**: Unlike traditional software, AI models are inherently complex and often function as "black boxes," which makes their decision-making processes opaque. For example,

most applications based on complex large language models (e.g., ChatGPT) produce outputs and reason over data in ways that are not fully transparent and understandable to human users. This lack of transparency poses significant risks, especially in high-stakes applications like healthcare or autonomous vehicles. Furthermore, users may struggle to trust systems they cannot fully understand or audit, which raises barriers to AI adoption.

- **Data protection issues due to dependency on large datasets**: AI systems rely heavily on vast amounts of data for their training and operation. This dependency creates vulnerabilities, such as data breaches, unauthorized access, or data poisoning attacks. Due to these vulnerabilities, malicious actors are likely to manipulate training datasets to corrupt AI outputs. It is, therefore, very challenging to ensure the integrity and security of these datasets.
- **Ethical concerns in decision making**: In many cases, AI systems operate in a biased way due to biases in their training data. Biased operations can lead to unfair or discriminatory outcomes. At the same time, the lack of accountability for decisions made by AI raises ethical concerns. For instance, it is not clear who is responsible when an autonomous system makes a mistake. These issues can cause public trust in AI technologies to collapse.

AI security concerns



Security measures implemented

Given the above-listed challenges, there is a need for robust frameworks for AI security and trust. For instance, such frameworks must foster AI transparency (e.g., based on AI explainability and AI interpretability techniques), while at the same time ensuring data security based on advanced encryption methods and secure data-sharing protocols that can protect sensitive information from breaches or misuse. Most importantly, it is important to promote ethical standards based on clear guidelines for fairness, accountability, and inclusivity that can mitigate biases and enhance public confidence in AI. The development of such a framework is a multistakeholder challenge involving policymakers, technologists, and

industry leaders. The latter must collaborate to develop strategies that address these challenges in ways that promote ethical AI use without hindering innovation.

5.2 AI SECURITY SOLUTIONS AND RISKS

5.2.1 Common Risk Factors

The above-listed security challenges of AI systems reflect a considerable number of AI-specific security risks, which must be embraced by modern security policies. Such risks include:

- **Adversarial attacks (e.g. evasion attacks)**: These are attacks that manipulate input data to deceive AI models in order to lead them to provide incorrect outputs or to compromise their integrity. For example, attackers can craft subtle changes to images or text in ways that cause misclassification in AI systems. Such attacks can compromise the reliability of critical applications such as autonomous driving and medical diagnostics.
- **Data poisoning**: Poisoning attacks take place during the training phase of the AI systems. They involve attackers that inject malicious data into datasets to corrupt the model's learning process. This can result in biased or harmful outputs, which can accordingly lead to wrong recommendations in decision-making processes like loan approvals.
- **Model theft and reverse engineering**: Proprietary AI models are also vulnerable to theft through unauthorized access or reverse engineering. In such cases, attackers replicate models by analyzing input-output patterns. This can lead to intellectual property theft.
- **Privacy breaches**: Many AI systems process sensitive data, which makes them targets for data breaches. There are adversarial actions like membership inference attacks that allow adversaries to determine if specific data was part of a model's training set. This leads to direct violations of the users' privacy.

"As I am under NDA, all I can say is that we have a rigorous 3 tiered process designed for engineers to understand and implement AI leaning into our current security system."
- Business leader survey respondent

- **AI-powered cyberattacks**: Nowadays, threat actors increasingly use AI to scale and optimize cyberattacks, such as phishing, ransomware, and denial-of-service (DoS) attacks. During the last couple of years, generative AI (GenAI) tools have significantly enhanced the sophistication of these attacks based on GenAI's ability to mimic human behavior and to create realistic fake content. As a prominent example, hackers have nowadays access to tools like FraudGPT and WormGPT, which are used to create and launch cyber-attacks via the Dark Web.

5.2.2 Mitigating AI Security Risks

Fortunately, AI vendors and security experts are offered effective solutions that can mitigate AI security risks. These solutions include:

- **Adversarial defense mechanisms**: These mechanisms hinge on the implementation of adversarial training processes, which expose models to deceptive inputs during their development time. Moreover, it is possible to use anomaly detection algorithms and input validation checks to identify and mitigate adversarial attempts.
- **Data protection measures**: It is also possible to ensure data integrity based on encryption for data at rest and in transit. In this direction, AI vendors and security solution providers can employ differential privacy techniques to anonymize datasets, while preserving their utility for training.
- **Robust model security**: To ensure the robustness of AI models, security professionals can employ techniques like access controls, secure deployment pipelines, and regular updates. They can also conduct penetration testing and vulnerability assessments, which helps them identify weaknesses proactively.
- **Continuous monitoring**: Another security measure involves the deployment of real-time monitoring tools to detect anomalous behavior in AI systems. In this direction, AI developers and deployers must conduct regular audits to ensure compliance with security standards and to help refine models over time.

"We have systems in place where we can easily update our systems ahead of any requirements while maintaining robust compliance frameworks."
- Business leader survey respondent

- **AI-driven cybersecurity tools**: AI is not just introducing new cyber-security risks. It is also a powerful tool for implementing security controls. In particular, it is possible to leverage AI techniques for threat detection. AI-based functionalities are typically integrated within conventional data-driven cyber-security tools, for example Security Information and Event Management (SIEM) systems, to enhance their threat detection and mitigation capabilities.
- **AI deployment at the edge computing:** In recent years, AI models and AI systems have been deployed at the edge (i.e., edge AI systems) rather than within a cloud computing infrastructure. For example, TinyML systems are deployed within IoT devices. Edge AI systems limit data transfer to remote datacenters, which can significantly reduce the attack surface of AI deployment. The latter is a key to alleviating vulnerabilities and reducing the likelihood of data breaches. The real-world security benefits of edge AI processing can be seen in applications like intelligent camera systems for monitoring elderly individuals in healthcare settings, where processing image and scene recognition directly on the device creates an inherently more secure system by keeping sensitive visual data local and eliminating risks associated with transmitting data to third parties for processing.

5.2.3 Hardware-Based Security Solutions

While many AI security solutions focus on software-level protections, hardware-based security features provide a fundamental layer of defense that is crucial for protecting AI systems. Modern processor architectures incorporate several key security technologies that help safeguard AI workloads:

- **Memory safety technologies**: Memory safety issues account for approximately 70% of all serious security vulnerabilities in computing systems, including AI applications. To address this, advanced processor architectures like Armv9 incorporate Memory Tagging Extension (MTE), which enables dynamic identification of both spatial and temporal

memory safety issues. This technology is particularly important for AI-based systems that process large amounts of data and complex model architectures. For example, Google has adopted Arm MTE in Android and committed to supporting MTE across the entire Android stack, providing enhanced security for millions of AI-enabled devices.

- **Secure virtualization**: As AI workloads increasingly run in virtualized environments, protecting the confidentiality and integrity of data becomes crucial. Realm Management Extensions, which form the basis of the Arm Confidential Compute Architecture, provide hardware-enforced isolation that secures data running in virtual machines from potential hypervisor compromises. This is especially critical in datacenters used for training advanced ML models, where multiple tenants may share computing resources. The same technology also helps secure edge computing systems where trained ML models are deployed.
- **Protection against code reuse attacks**: Modern AI systems face sophisticated attacks that can repurpose existing code for malicious purposes. Technologies like Arm Pointer Authentication (PAC) and Branch Target Identification (BTI) provide robust protection against code reuse attacks such as return-oriented programming (ROP) and jump-oriented programming (JOP). These protections are particularly important as attackers increasingly use AI tools to develop more sophisticated attack methods. These security features are being deployed across both high-performance application processors and microcontrollers used in IoT devi.
- **Standardized security framework**: Beyond individual security features, industry-led security frameworks like PSA Certified provide a comprehensive approach to device security. This framework establishes security best practices and certification processes that help ensure AI-enabled devices meet robust security standards from the silicon level up. By adhering to these standards, manufacturers can demonstrate their commitment to security while providing customers with verifiable security assurances.

These hardware-based security solutions complement software-level protections to create a comprehensive security architecture for AI systems. As AI workloads become more prevalent across different computing environments – from cloud datacenters to edge devices– the role of hardware security becomes increasingly critical in protecting sensitive AI models and data.

5.3 DATA PROTECTION IN AI APPLICATIONS

5.3.1 Data Protection Risks

As already outlined, AI systems require large, high-quality datasets to function properly and efficiently. In several cases, these datasets include personal or sensitive information, which introduces significant risks. For instance, sensitive data (e.g., health records and financial information) can be exposed during storage, transmission, or processing. This can lead to privacy breaches and potential misuse. Overall, AI systems face several key challenges when it comes to safeguarding data:

- **Data breaches**: Large datasets are attractive targets for cyberattacks. Unauthorized access can compromise personal information, which can subsequently lead to identity theft or financial fraud.
- **Lack of transparency**: In many cases, AI users lack information and clarity about how AI data are collected, stored, and used.
- **Bias and misuse**: Improper handling of data can perpetuate biases or lead to discriminatory outcomes. This is a setback to ensuring ethical AI deployments.

5.3.2 Solutions for Enhancing Data Protection

The following technologies, strategies, and measures can be used to address the above-listed data protection challenges:

- **Encryption techniques**: AI vendors and system developers offer advanced data encryption methods, allowing them to safeguard sensitive data against data breaches. For instance, techniques like

homomorphic encryption allow computations on encrypted data without decryption, which boosts privacy and data protection. As another example, differential privacy techniques add statistical noise to datasets in order to protect individual identities while maintaining analytical utility.

- **Secure multiparty computation (SMPC)**: SMPC is a novel technique that enables collaborative model training without exposing raw data. In the scope of this technique, each party processes encrypted inputs locally, which ensures that sensitive information remains secure.
- **Regulatory compliance**: AI applications are subject to stringent data protection regulations, which safeguard data privacy. As a prominent example, frameworks like the European General Data Protection Regulation (GDPR) help ensure that data is collected, processed, and stored ethically. In this direction, AI data processors and AI system deployers can employ techniques like data minimization and anonymization. Furthermore, they can also conduct Data Protection Impact Assessments (DPIAs) in order to identify how they can stay compliant with applicable laws and regulations.

5.4 CHALLENGES AND SOLUTIONS FOR TRUSTWORTHY AI SYSTEMS

5.4.1 Trustworthiness Challenges

AI systems face significant trustworthiness challenges that typically stem from their lack of transparency, susceptibility to bias, and difficulties in verifying the reliability of their outputs. These issues can be very critical in high-stake AI applications, such as healthcare, finance, and criminal justice. More specifically, the main AI trustworthiness challenges are as follows:

- **AI black boxes**: Many AI systems operate as "black boxes." This means that their decision-making processes are not clear and transparent to their users. There are many cases where these processes are not transparent even to their developers. This lack of interpretability makes it

difficult to understand how models arrive at specific conclusions, which hinders accountability and trust. For example, an AI system for medical screening might suggest a specific diagnosis, yet it can hardly explain the rationale behind its decision.

- **Algorithmic bias**: Bias is a pervasive issue in AI, which is often introduced based on skewed or limited training data. The latter can result in discriminatory outcomes, such as denying loans to certain demographic groups or unfair hiring practices. Even with efforts to mitigate bias, algorithmic bias can persist due to the inherent complexities of real-world data and due to the varying definitions of fairness across contexts.

  Reliability verification: Ensuring consistent and reliable outputs from AI systems is not a trivial task. Hence, models may perform well during testing but fail in real-world scenarios due to unforeseen variables or data shifts. This unpredictability undermines user confidence and raises concerns about the robustness of AI applications.

5.4.2 Strategies to Enhance Trustworthiness

The following measures and techniques can be nowadays employed to safeguard AI trustworthiness:

- **Security by design**: A security-by-design approach integrates security considerations from the hardware level up. This approach involves building security features directly into the processor architecture and silicon, rather than trying to add security measures after deployment. For example, hardware-based security features can help provide robust protection against memory-related vulnerabilities, secure the execution of AI workloads in virtualized environments, and protect against sophisticated code manipulation attacks.
- **Explainable AI (XAI)**: XAI is a special research direction in AI, which aims at making the operations and outcomes of AI models transparent and understandable by users. For instance, techniques like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic

Explanations are nowadays widely used to explain the internal operations of ML models, which fosters trust in AI systems. As another example, counterfactual techniques are also used to provide "what-if" insights on how AI model inputs influence outputs. This is one more approach to reinforcing trust in the operations of AI systems.

- **Bias detection and mitigation**: It is also possible to address bias issues proactively, i.e. during the AI data preparation and model training process. In this direction, fairness-aware algorithms, diverse data sourcing, and regular audits can be employed to identify and reduce biases. Furthermore, organizations are prompted to adhere to standards and regulations like the AI Act of the European Union The latter comprise fairness detection and mitigation mandates, which are especially applicable in the case of AI issues.
- **Rigorous testing and validation**: Systematic testing across diverse scenarios is one more effective way of verifying the reliability of an AI system. This includes stress-testing models under varying conditions, monitoring them for performance drift over time, and conducting independent audits to ensure adherence to ethical guidelines.
- **Independent security certification**: Such frameworks play a crucial role in verifying and validating these security implementations. These frameworks provide structured evaluation processes that test products against defined security requirements and help build trust by providing independent verification of security claims. It also ensures compliance with evolving security standards and regulations.

5.5 EVALUATION METRICS FOR SECURITY AND TRUST

To evaluate the security and trustworthiness of AI systems, organizations require robust metrics that address their key vulnerabilities and ethical concerns. These metrics can be used to help ensure these systems are reliable and fair and preserve privacy when required. These metrics provide

a framework for assessing AI model performance under various conditions and adherence to ethical standards.

Some of the most prominent security and trust metrics for AI systems include:

- **Robustness against adversarial attacks**: Adversarial robustness measures an AI model's ability to withstand maliciously crafted inputs designed to deceive it. Relevant metrics for AI resilience include:
  (i) Attack success rate.;
  (ii) Robustness radius, i.e., the maximum perturbation a model can tolerate.
  (iii) Adversarial accuracy, i.e., the percentage of correct predictions on adversarial examples. For instance, a high robustness score indicates that the model is less susceptible to adversarial manipulation, which is very important in the case of applications in security-sensitive domains like finance or healthcare.
- **Resilience to data poisoning**: Data poisoning resilience evaluates how well a model performs when its training data is tampered with. Relevant metrics include accuracy degradation under attack and the detection rate of poisoned samples. Models with strong defenses (e.g., anomaly detection, robust training) demonstrate minimal performance loss even when exposed to malicious data.
- **Interpretability scores**: Many (XAI) metrics assess how well a model's decisions can be understood by humans. In several cases, these measures are combined with user satisfaction indicators, which gauge the user friendliness of the AI explanations as well.
- **Fairness metrics**: Fairness metrics like demographic parity and equal opportunity measure whether AI systems treat all demographic groups equitably. These metrics help identify and mitigate biases while helping to ensure that models do not perpetuate discrimination.
- **Differential privacy guarantees**: Differential privacy quantifies the privacy risk associated with individual data points in a dataset. It comprises a "privacy budget" parameter, which is the trade-off between

privacy and utility. Lower “privacy budget” values indicate stronger privacy guarantees.

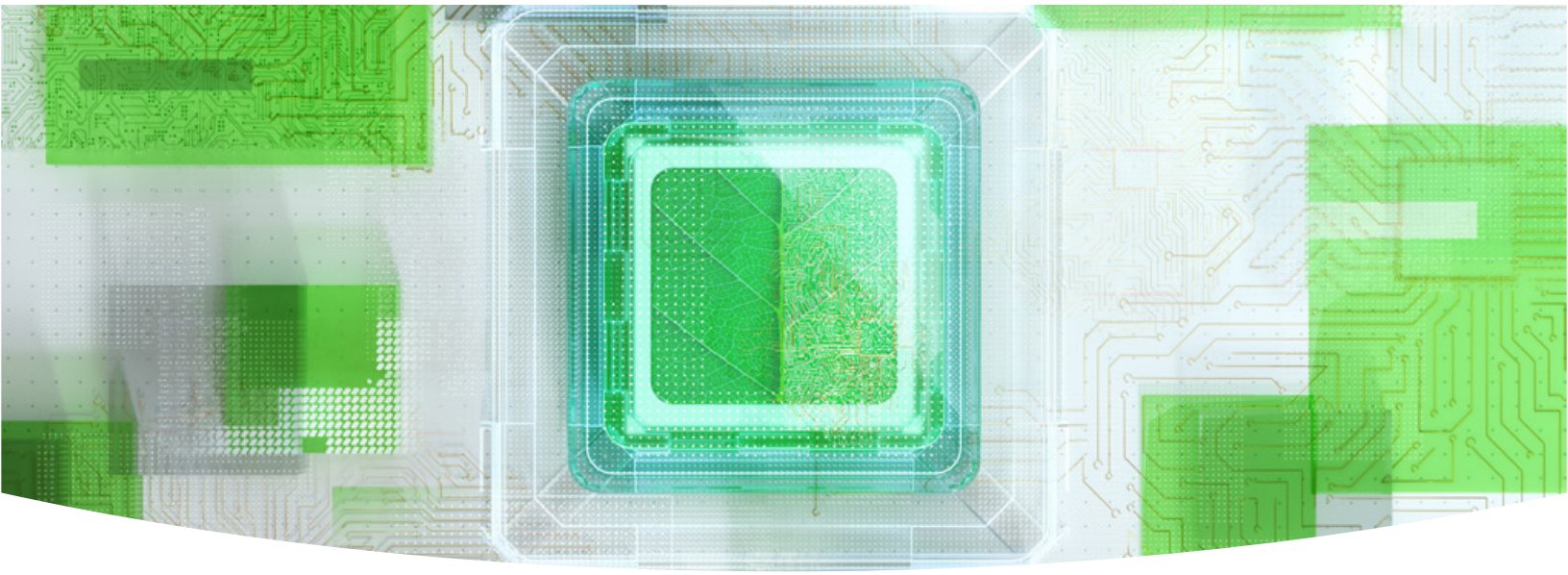
– **Data leakage risk assessments**: These metrics evaluate the likelihood of sensitive information being inferred from model outputs or training datasets. The latter ensures that private data remains secure during use.

CONCLUSION

Overall, AI systems vendors and deployers must address the challenges of security, trust, and privacy. As AI systems become increasingly integrated into critical aspects of society, it is very important to ensure their robustness, fairness, and ethical use. The latter is essential for fostering public confidence and maximizing their benefits.

Interdisciplinary collaboration plays a pivotal role when it comes to addressing AI security and trust challenges in production environments. Technologists must work alongside policymakers, ethicists, and industry leaders to establish robust frameworks that balance innovation with accountability. At the same time, there is a need for continuous advancements in security measures (e.g. XAI techniques, privacy-preserving technologies, and fairness-aware algorithms) to keep pace with evolving threats and societal expectations. As a call to action, stakeholders across sectors must prioritize trust and security as foundational principles for sustainable AI development.

These values must be embedded into the design, implementation, and governance of AI systems to help ensure that emerging AI technologies serve as a force for good. The future of AI must empower individuals, drive innovation, and address global challenges, while at the same time safeguarding ethical standards. The path forward requires collective effort, but the rewards of trustworthy and secure AI are certainly well worth the investment.



CHAPTER 6

Sustainability as a Core Metric for AI Readiness

DR VANESSA JUST,
FOUNDER AND CEO OF JUS.TECH GMBH

DR NICOLE HÖHER,
PROJECT MANAGER FOR SUSTAINABILITY & DIGITALISATION AT JUS.TECH GMBH

AI has emerged as a transformative force, revolutionizing industries, and enhancing efficiencies across different sectors of the economy. However, this rapid advancement comes with environmental considerations; the energy consumption associated with AI technologies poses a challenge to global environmental sustainability efforts. At the same time, AI offers tools to combat climate change, optimize energy use, and enhance environmental monitoring. As such, AI has a dual role as both a substantial new source of energy demand and a potential ally in promoting sustainability, exploring strategies to balance technological progress with environmental stewardship.

6.1 AI'S RISING ENERGY CONSUMPTION AND ITS IMPLICATIONS

The rapid expansion of AI applications has driven an unprecedented surge in computational demand, particularly for training complex models, such as large language models. These AI systems require immense processing

power, resulting in a significant increase in energy consumption, not to mention significant amounts of water for cooling. This growing demand has far-reaching implications for global electricity usage and carbon emissions.

Today's datacenters already consume lots of power: Globally 460 terawatt-hours (TWh) of electricity are needed annually. That's equivalent to the entire country of Germany. In the United States, datacenter electricity consumption was 2.5% of the U.S. total (~130 TWh) in 2022 and is expected to triple to 7.5% (~390 TWh) by 2030, according to the Boston Consulting Group. That's the equivalent of the electricity used by about 40 million U.S. houses – almost a third of the total homes in the U.S.

"Without proactive measures, AI-driven energy consumption could push the world further off track from climate targets."

This escalating energy consumption presents a critical challenge to electrical grid capacity and reliability worldwide. Many power grids, already under strain from increasing electrification and growing populations, must now accommodate the rapid expansion of AI-driven workloads. The transition to renewable energy and its inherent intermittency also brings its own challenges to balancing grid supply and demand. Without significant infrastructure investment, the risk of grid instability and supply constraints grows.

Utilities and energy providers are already grappling with the impact of AI's rising energy consumption. In several regions, notably Virginia in the United States and Ireland, high-voltage transmission networks are becoming congested, slowing the deployment of new datacenters. To manage these limitations, utility companies in some areas have already implemented moratoriums or rationing policies on new datacenter connections.

Addressing these challenges requires a multipronged strategy that includes expanding grid capacity, improving transmission infrastructure, integrating renewable energy sources, and adopting energy-efficient AI technologies. Without coordinated efforts between governments, industry

leaders, and energy providers, AI's unchecked energy demand could strain power systems and undermine broader environmental sustainability goals.

6.2 ENHANCING ENERGY EFFICIENCY IN EDGE AI

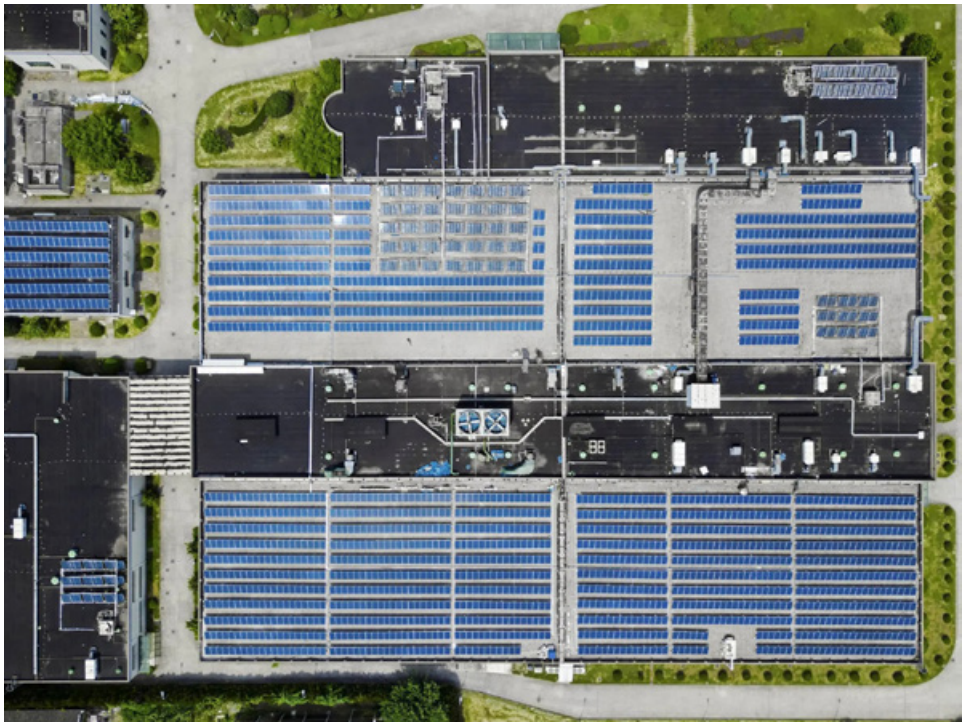
AI's environmental footprint extends beyond large-scale datacenters, reaching decentralized environments such as edge computing and IoT devices. As AI adoption accelerates, these technologies are playing an increasingly important role in processing data closer to the source, reducing latency, bandwidth consumption, and reliance on cloud infrastructure. This shift offers environmental benefits for AI inference but also comes with its own environmental challenges; edge and IoT devices require computational power, and as their numbers grow exponentially, so does their cumulative energy demand. Addressing this challenge to realize the many benefits is essential for ensuring a sustainable AI ecosystem.

- **Energy-Intensive Edge Computing**. Edge computing enables real-time data processing without transmitting large datasets to the cloud. While this helps reduce network congestion and improve efficiency, the widespread deployment of AI-powered edge devices leads to increased energy consumption. Without optimized architectures, the sheer volume of edge devices could undermine sustainability efforts rather than enhance them.
- **IoT Devices and Inefficient Power Usage**. The rapid adoption of IoT devices has revolutionized industries by enabling real-time monitoring, automation, and decision-making. IoT systems may operate inefficiently, consuming excessive energy due to outdated hardware and software architectures. While integrating AI can help improve efficiency and predictive analytics, balancing performance with sustainability remains a challenge.

To mitigate the energy impact of AI in decentralized environments, a comprehensive, multilayered approach is needed, focusing on hardware innovation, algorithmic efficiency, and system-wide energy management.

- **Smarter AI Hardware (*Hardware Optimization*).** Using specialized low-power chips can cut energy use while keeping AI systems fast and effective. For example, AI accelerators are designed to use less power while still processing AI tasks quickly; and efficient microcontrollers and smart chips help AI devices run without draining too much electricity.
- **Making AI Models More Efficient *(Algorithmic efficiency).*** AI models can be right-sized for certain use cases so they use less computing power while still being accurate. Some ways to do this include simplifying AI models (pruning & quantization); cutting out unnecessary calculations so AI runs faster and with less energy; and smaller, lightweight AI models, built in a way that doesn't require heavy processing, making it easier to run on phones, smart home devices, and other low-power devices.
- **Smarter Energy Management in AI Systems *(System-Level Strategies)***. AI systems can improve their energy use management by adjusting power based on what's needed at any moment. Methods such as dynamic power adjustment allow AI devices to increase or decrease their processing power depending on the task, to help prevent waste. In addition, balancing workloads between devices allows AI tasks to be shared between the cloud, edge devices, and IoT gadgets to help ensure no single system is using more energy than necessary.

By making hardware, AI models, and power management smarter, AI can be just as powerful while using less energy, helping the environment, and reducing costs.



6.3 FOCUS ON INNOVATIONS IN ENERGY-EFFICIENT AI HARDWARE

While software optimizations and intelligent workload management can enhance AI efficiency, hardware innovation remains a foundational pillar in AI's sustainability journey. The semiconductor industry plays a pivotal role in developing low-power, high-performance processors that support AI workloads without excessive energy usage.

Companies like Arm are leading the charge in designing energy-efficient processors that can minimize AI's environmental impact, while maintaining high computational performance. Some of the most significant advances include:

- AWS Graviton processors: Built on Arm technology, these processors help reduce workload carbon intensity by up to 67% compared to traditional x86 chips, enabling more efficient cloud-based AI workloads.
- Mobile & edge AI acceleration: Arm-based accelerators in mobile and

edge computing lower energy consumption by 50% to 80% compared to general-purpose GPUs, making AI inferencing on smartphones, IoT devices, and autonomous systems far more efficient.

- Local processing: Arm-powered on-device AI processing eliminates the need for constant cloud data transmission, helping to reduce network energy usage while enhancing real-time decision-making.
- Datacenter optimization: Arm Neoverse CPUs power some of the most energy-efficient cloud infrastructures, cutting server rack energy consumption by up to 40% and potentially making hyperscale AI deployments more sustainable.

AI's expansion into edge computing and IoT ecosystems presents both an opportunity and a challenge. While these technologies enhance real-time processing and help reduce cloud dependency, their growing energy footprint must be actively managed.

By integrating energy-efficient hardware, optimizing AI algorithms, and implementing intelligent system-wide energy strategies, the industry can significantly reduce power consumption without sacrificing AI's potential for innovation.

6.4 AI AS A CATALYST FOR GLOBAL CLIMATE SOLUTIONS

Despite its high energy demands, AI has the potential to be a game-changer in tackling global environmental challenges by improving efficiency, reducing waste, and accelerating the renewable energy transition. One of its most promising applications is in renewable energy optimization, particularly in managing wind and solar power.

AI-powered wind energy forecasting helps predict wind patterns with greater accuracy, allowing wind farms to operate more efficiently by optimizing

turbine positioning and energy storage. Countries where wind energy plays a major role in the energy mix, can benefit significantly from these advances, ensuring a more stable and predictable supply of clean energy.

Similarly, solar energy management is being revolutionized by machine learning algorithms that analyze sunlight exposure, weather conditions, and energy demand to optimize the placement, maintenance, and efficiency of solar panels. Regions with high solar potential, such as California, the Middle East, and Australia, can leverage AI to maximize solar energy generation, reducing dependency on fossil fuels and increasing grid resilience.

Beyond renewable energy, AI is transforming climate modeling and environmental monitoring, providing more precise tools for scientists and policymakers. AI-driven climate models help climate scientists, meteorologists, and corporate risk assessors to better understand extreme weather patterns, predict climate-related disasters, and develop more effective mitigation strategies. These advances are critical for countries that are on the frontlines of climate change impact and require more reliable forecasting tools to prepare for and adapt to natural disasters.

AI also enhances environmental monitoring, helping scientists track deforestation, biodiversity loss, and ocean health. Organizations like the World Resources Institute (WRI) and the United Nations Environment Programme (UNEP) are already deploying AI-powered satellite imagery analysis to detect illegal deforestation in the Amazon and to monitor marine ecosystems affected by climate change.

Another crucial area where AI is making an impact is energy efficiency in transportation. AI is revolutionizing autonomous transportation, optimizing traffic flow, improving public transit systems, and reducing overall fuel consumption through intelligent routing and vehicle efficiency models.

Forward-thinking cities are integrating AI-driven transport systems to help reduce congestion and lower carbon emissions and air pollution, showcasing how urban centers can benefit from smarter, AI-powered mobility solutions.

By combining machine learning, big data, and real-time analytics, AI is not just a high-energy consumer, it is also a critical enabler of sustainable solutions. Whether through optimizing renewables, refining climate forecasting, or enabling smart transport systems, AI has the potential to reshape how we combat climate change and transition to a more sustainable future. The key lies in ensuring that AI's own energy footprint is minimized so it can be an effective force for environmental progress.

6.5 COLLABORATION FOR SUSTAINABLE AI

Tackling the environmental impact of AI requires a unified, multistakeholder approach that brings together governments, industry leaders, research institutions, and civil society organizations (CSOs). No single entity can address these challenges alone, meaningful progress demands collaboration at a global scale. Recent initiatives, such as the AI Action Summit in Paris, have highlighted the need for coordinated efforts to ensure that AI development aligns with long-term environmental sustainability goals. The Paris summit brought together participants from over 100 countries, including government leaders, international organizations, CSOs, the private sector, and academic communities, and aimed to foster discussions on aligning AI development with global sustainability goals and promoting responsible AI that supports environmental policies.

From a climate perspective, a notable outcome is the launch of the 'Coalition for Sustainable AI' initiated by France, in collaboration with the UNEP and the International Telecommunication Union (ITU). This coalition

seeks to build a global community dedicated to aligning AI development with environmental objectives, fostering responsible AI that supports sustainability policies.

Cross-sector collaboration plays a critical role in steering AI development towards sustainability. Only by uniting key stakeholders such as governments, international organizations, the private sector, and CSOs can a comprehensive approach to addressing the environmental and ethical implications of AI be found. Such collaborative efforts are essential to help ensure that AI technologies contribute positively to global sustainability goals, while mitigating potential adverse impacts.

6.6 BALANCING INNOVATION AND RESPONSIBILITY

AI represents a paradox, an energy-intensive technology with immense potential to deliver positive environmental sustainability impacts. The challenge now is to ensure that its deployment does not undermine the very progress it seeks to enable. The industry must strike a balance between AI's exponential growth and the urgent need for energy-efficient innovation.

By designing inference-based AI systems that are power and energy efficient, shifting toward decentralized and edge computing, and integrating AI with renewable energy infrastructure, we can minimize its environmental impact while maximizing its benefits. The decisions made today are shaping the trajectory of AI's role in the global sustainability movement. It is no longer a question of whether AI can be sustainable, but whether we will take the necessary steps to make it so.

ARM SIDEBAR

Sustainable AI: Balancing Innovation with Environmental Impact

Maureen McDonagh, Head of Sustainability, Arm

As AI continues to transform industries and society, the urgency to balance its rapid growth with environmental responsibility has become a key consideration. While AI undoubtedly drives immense energy and power demands, it can also be used as a tool for tackling wider sustainability challenges.

Therefore, to fully harness AI's potential and mitigate its potential drawbacks, the industry must adopt sustainable practices across development, deployment, and usage, while exploring ways to utilize it effectively to mitigate environmental impact.

The Environmental Impact of AI

The growth of AI is fueling a surge in energy demand, particularly in datacenters that power its training and inference processes. Global electricity consumption by datacenters is projected to triple by 2030,

accounting for 3% of worldwide power use, according to a report by Deloitte. A key contributor to this growth is going to be AI and its applications.

Training large-scale models involves vast amounts of computational power. Meanwhile, inference – where these models are applied to new data – are estimated to be ten times more energy-intensive than traditional computational queries.

This greater demand for energy raises costs and increases greenhouse gas (GHG) emissions, which exacerbates the effects of climate change. Without proactive measures, AI-driven energy consumption could push the world further off track from climate targets, with projections indicating an increase of over 2°C in global temperatures; breaching the recommendations to limit the rise to a much safer 1.5°ct.

However, this increased environmental impact is not just applicable to carbon emissions. New modelling shows that generative AI could lead to a 1,000-fold increase in electronic waste by 2030 without effective waste reduction strategies. In fact, a study by The Register, an online enterprise technology news publication, reveals that e-waste may rise by up to 2.5 million tons each year by 2030 if effective reduction measures are not introduced.

Sustainable AI Initiatives and Technologies

To minimize AI's environmental footprint, the industry must embed sustainability into every layer of AI systems, from hardware to software, while supporting the broader energy transition.

Efficiency can be incorporated across all levels of the stack, from the foundational hardware to the software. These cover the following areas:

- CPU and GPU designs that can be optimized for AI and power-efficiency;
- Innovative silicon designs utilizing 3D chip architectures and memory hierarchies to minimize data movement and energy consumption;
- Exploring innovative materials to replace those with a high environmental impact;
- Innovations in edge computing enabling high-performance, power-efficient AI solutions;
- Efficient software designs where concepts, such as software carbon intensity, are utilized to ensure sustainability considerations are embedded from the start.

Beyond the stack, there are a number of initiatives that can support more sustainable AI.

Firstly, moving the processing of AI inference workloads to edge devices to help minimize the energy costs associated with data transmission to the cloud. Processing AI workloads closer to the data source, such as on local devices or servers, reduces latency, energy transmission losses, and even improves privacy by limiting the need to transmit sensitive information. To make edge AI processing easier, techniques like model compression and pruning can also be applied to reduce computational loads without sacrificing performance.

Secondly, there are ways to optimize AI training workloads to limit their environmental impact. For example, AI training can be performed in regions with abundant low carbon or renewable energy.

Finally, there are various green coding practices that can be implemented, such as writing optimized, resource-efficient code that can reduce computational demands.

The Role of AI in Climate Solutions

AI itself is a powerful enabler of sustainable practices, offering solutions for energy optimization, climate adaptation, and emissions reductions, with this being especially relevant as part of the wider global low carbon transition.

AI improves forecasting for renewable energy resources, like solar and wind, enabling optimization of grid operations, and enhancing energy storage performance. It also supports grid efficiency by predicting peak demand and reducing energy losses.

From a climate change adaptation perspective, AI can improve climate modelling and is already providing early-warning systems for natural disasters. For example, the UNICEF Arm-powered AI-powered flood modeling in Malawi demonstrates how AI can mitigate climate risks, saving lives and reduce recovery times.

Finally, AI-driven innovations can reduce waste and emissions across critical sectors by optimizing supply chains and enhancing transportation systems.

Collaborating for a Sustainable AI Future

A sustainable AI future requires coordinated efforts across government, industry, and academia. Key efforts driving sustainable AI development include:

- **Policy and regulation**: The EU AI Act demonstrates how regulatory frameworks can guide responsible AI development by emphasizing transparency, accountability, and ethical practices.
- **Partnerships for change**: Organizations like Arm, in collaboration with UN bodies and other stakeholders, are driving initiatives for climate-focused AI solutions.

– **Sustainable semiconductor development**: The U.S. NIST launched an initiative with $100 million in funding to use AI for creating sustainable semiconductor materials within five years, encouraging academia-industry collaboration.
– **Standardization**: Industry-wide efficiency metrics, such as those spearheaded by policymakers, can ensure AI aligns with sustainability goals. In fact, the world's largest technology companies, including AWS, Microsoft, and Google, are advocating for Environmental Product Declarations to assess and potentially reduce embodied emissions in datacenter infrastructure.

Leading AI Responsibly

The AI industry has a unique opportunity and responsibility to lead by example. By prioritizing sustainability, it can demonstrate that technological innovation and environmental stewardship are not mutually exclusive. Key actions the sector can take include:

– Investing in renewable energy to power AI operations;
– Innovating for power-efficient hardware and software;
– Championing green AI practices that align with global climate goals.

This approach helps ensure that AI's transformative potential is harnessed responsibly, enabling progress that benefits both humanity and the planet. Through bold commitments and collaborative action, the industry can define a future where AI accelerates solutions to the very challenges it contributes to today.

CHAPTER 7

Building an AI-Ready Culture

MARK HINKLE,
CEO AND FOUNDER OF PERIPETY LABS, FOUNDING PUBLISHER THE ARTIFICIALLY INTELLIGENT ENTERPRISE

AI adoption in companies is accelerating, but employees on the ground are struggling to keep up. There's often a palpable gap between the rapid pace of AI advances and what employees feel prepared to handle.

The truth is that many workers feel overwhelmed by constantly evolving tools and lack the training to use them effectively. Nearly seven in ten employees never use AI at all in their work, and only about one in ten use AI every week. One major reason for this usage gap is that employees haven't been given enough guidance or training – only 15% of U.S. workers say their organization has communicated a clear AI strategy, and a mere 11% feel "very prepared" to work with AI in their role.

When people aren't sure why or how to use new AI tools, it's no surprise they stick to familiar routines. This lack of preparedness is widespread. Even though 75% of companies have now adopted some form of AI technology, just one-third of employees have received any AI-related training in the last year. In other words, companies are rolling out AI faster than their teams can absorb it. The result? Employees feel left behind and

"We organize week-long trainings. Provide comprehensive online courses to help employees fully grasp the concept of AI development."
- Business leader survey respondent

anxious. They might even quietly push back against AI projects – skipping the use of a new system or voicing doubts – not because they hate innovation but because they feel unready and unsupported.

Corporate leaders risk underutilizing their AI investments when the workforce isn't on board. Consider that only about one in five digital transformation initiatives fully achieves its goals, often due to cultural and adoption issues. When employees aren't empowered to use AI, those expensive platforms and algorithms can turn into costly shelfware, delivering only a fraction of the promised productivity gains. In short, technology alone can't transform your business; your people have to transform along with it.

As we saw in the AI Readiness Index survey results (Chapter 1), enterprises recognize they have significant work to do when it comes to building AI-ready organizations. The gap between AI adoption intentions and organizational readiness presents both a challenge and an opportunity for business leaders looking to harness AI's transformative potential.

- Only 39% of leaders report their organization has a clearly defined comprehensive AI strategy. The remaining organizations are either still developing their approach (13%) or have no formal strategy at all (4%).
- Over a third (34%) of respondents indicated their organizations are either significantly under-resourced or under-resourced with AI talent. Even more concerning, 39% reported having no dedicated programs for developing AI skills among employees.
- While 80% of business leaders say there is a designated budget for AI initiatives, only 9% allocate more than 20% of their technology budget to AI, suggesting most organizations are still taking a cautious approach to AI investment.
- Improving operational efficiency (80%) and enhancing customer experience (70%) top the list of AI initiatives.

"Selected personnel from each department undergo intensive Python development education twice a week for one month, with certification exams conducted upon completion."
- Business leader survey respondent

- 48% of leaders worried about data privacy breaches through model extraction.
- Other top challenges include lack of skilled talent (49%), data quality and accessibility issues (46%), and regulatory and compliance concerns (46%).

7.1 BUILDING AN AI-READY CULTURE: A ROADMAP

So, how can you turn things around and make your organization truly AI-ready? The good news is that the very obstacles we've described – the skills gap, employee anxiety, and underutilized tools – can be overcome with the right strategies. It starts with a shift in culture and mindset from the C-suite to the front lines.

Leaders must actively bridge the gap through education, communication, and support. Leaders must bridge the AI skills gap to alleviate employee frustration and resistance. In practice, building an AI-ready culture means investing as much in your people as you do in technology. Here's how to get started:

7.2 INVEST IN UPSKILLING AND CONTINUOUS LEARNING

Make AI literacy a core competency across your organization. Provide training programs, workshops, and easy-to-access online courses to help employees at all levels build confidence with AI tools. This isn't a one-time effort – encourage continuous learning as AI evolves.

Not only will this boost adoption, but it also pays off financially: developing your existing talent is far more cost-effective than hiring new experts (hiring a new worker can cost up to seven times more than upskilling an existing

"Our organization offers a comprehensive AI training program that includes workshops, online courses, and practical projects."
- Business leader survey respondent

employee). Fortunately, your employees want to learn – 57% of workers say they want their company to provide AI training to grow their skills. By offering meaningful upskilling opportunities, you signal that AI is a priority and that every team member has a stake in this transformation.

While internal training programs are essential, organizations can also benefit from the broader educational ecosystem discussed in Khaled Benkrid's sidebar to this chapter, including academic partnerships and industry consortiums like the Semiconductor Education Alliance.

7.3 LEAD WITH CHANGE MANAGEMENT AND CLEAR COMMUNICATION

Adopting AI is as much about people and processes as it is about tech. Combat resistance by involving employees early and communicating openly about your AI vision. Set a clear AI strategy and share it widely – what are you implementing and why? How will it make jobs easier or more interesting? Helping people understand the purpose behind the technology, can calms AI uncertainty and reduce fear.

It's also crucial to address the elephant in the room: job security. Reassure your teams that AI is there to augment their work, not replace them, for example by highlighting how automating tedious tasks frees up time for more creative, high-value work. Back up these assurances with action: offer reassignment or additional training for roles that are likely to change.

Also, empower leadership and managers to champion the change. When top executives and team leaders actively use AI tools and encourage their teams to try them, it sets the tone that AI is a trusted part of the workflow, not a fad. For instance, JPMorgan Chase recognized the importance of top-down support; one of its senior leaders announced that every new

employee would receive prompt engineering training to get them ready for an AI-driven future.

By making AI knowledge an onboarding requirement, the company send a powerful message that embracing AI is a core value. Effective change management also means creating feedback loops – giving employees a voice to share their concerns and insights as you implement AI. This helps you identify stumbling blocks early and helps staff feel heard during the transition.

7.4 FOSTER A HANDS-ON, INCLUSIVE AI CULTURE

To truly embed AI into your business DNA, encourage a culture of experimentation and inclusion. Make AI tools available to employees in their day-to-day work and encourage them to "play" with these tools in a low-pressure setting.

One idea is to designate team-based AI champions or mentors – tech-savvy staff who can help their peers discover useful AI use cases and troubleshoot issues. Celebrate quick wins and share success stories: for example, if someone on the sales team uses an AI tool to automate part of a proposal and saves hours, highlight that achievement in your internal newsletter or meetings.

These stories make AI's benefits concrete and relatable, turning skeptics into curious adopters. It's also wise to integrate AI goals into performance metrics or incentives, gently nudging teams to incorporate AI where it makes sense.

Remember to be inclusive in your approach: ensure training and tool access reach everyone, not just tech teams or a select few. Today, many

organizations have a bias where only certain roles get AI training – but success with AI requires broad participation.)One thing I strongly suggest avoiding is an AI-only group or a Tiger Team. If only a small group gets skilled up, you risk creating an AI "have and have-not" divide. Instead, strive for AI fluency across departments. This inclusive approach was echoed in a recent global workforce study, which found that empowering workers at all levels is key to unlocking AI's full potential. When everyone, from entry-level employees to senior managers, has at least a baseline understanding of AI, your company can truly capitalize on collective intelligence and creativity augmented by these new tools.

AI Adoption isn't Just a Digital Transformation – it's a People Transformation.

By following this roadmap – upskilling your people, managing change thoughtfully, and nurturing an innovative culture – you create an environment where AI can thrive. It transforms AI from something intimidating into something empowering. Employees who once felt overwhelmed or threatened by AI start to feel curious and competent, especially as they see their skills grow and their workload shift to more rewarding tasks. And the impact on the business is tangible.

Companies that invest in their workforce's AI readiness stand to maximize the productivity gains of their AI investments. Instead of half-used software

licenses or stalled pilot projects, you have teams actively leveraging AI to solve problems, make better decisions, and drive efficiency. In short, an AI-ready culture turns AI from a shiny new thing in the toolbelt into a natural extension of how work gets done.

Close the skills gap by empowering your employees, and watch resistance melt into enthusiasm. With the right culture in place, your organization can ride the AI wave with confidence, ensuring that both your people and your technology are working in sync to move the business forward.
The companies that get this right not only see higher ROI on their AI projects but also a more engaged, future-ready workforce poised to innovate. In the age of AI, your competitive advantage boils down to your team's ability to adapt and excel alongside intelligent machines. By building an AI-ready culture today, you prepare your business – and your people – to thrive tomorrow.

ARM SIDEBAR

Addressing the Skills Challenges in an AI-Driven Workforce

Khaled Benkrid, Senior Director, Education and Research, Arm

In a world where humans and machines are working together more than ever, the ability to build and use AI tools effectively is becoming a fundamental skill. Yet, as AI continues to evolve at an unprecedented pace, workforce development is struggling to keep up. Many professionals lack the specialized knowledge needed to develop, deploy, and optimize AI-driven solutions, creating a widening skills gap that threatens to stall innovation.

Current Skills Challenges in the AI Workforce

As AI adoption accelerates across industries, the demand for a workforce proficient in AI technologies has surged. Emerging technologies, AI tools and systems require a level of familiarity and expertise that many workers lack, making it difficult for businesses to find professionals with the expertise needed to develop and integrate AI effectively.

This skills gap can affect every job role, making staff and professional development essential. Companies must invest in training programs that

help employees understand AI's capabilities and applications. Furthermore, the future workforce must combine human ingenuity with new and emerging AI technologies; going beyond just the technical skills.

"In a world where humans and machines are working together more than ever, the ability to build and use AI tools effectively is becoming a fundamental skill."

Addressing these challenges requires engineers, practitioners, policy makers and students to acquire foundational skills and concepts through education and training. To meet this demand, the current curricula must undergo substantial revision, equipping students with knowledge that aligns with new AI-driven programming and decision-making.

Context-specific Education for Regional and Global Variations in Skills Challenges

As AI is being adopted worldwide, the skills gap has emerged as a widespread challenge. There are notable regional variations in readiness and focus. While some countries lead in AI adoption, regulation, and standard-setting, others lag behind. Even within countries, priorities differ based on regional needs. For instance, in the U.S., rural communities may emphasize AI applications in agriculture as opposed to more urban communities that will focus on AI for smart cities–highlighting the importance of context-specific education and training.

The Evolution of Educational Requirements

The emergence of AI has disrupted traditional educational requirements and degree programs. While computing has been a relatively recent addition to school curricula, it is already being transformed by AI. This demands regular updates to ensure that education keeps pace with technology advancementss. As a result, learning models must become more agile to let the dynamic curriculum evolve alongside industry needs.

"Companies must invest in training programs that help employees understand AI's capabilities and applications. Furthermore, the future workforce will need to combine human ingenuity with new and emerging AI technologies; going beyond just the technical skills."

Reskilling efforts must begin at the pre-university level, focusing on foundational abilities, knowledge and skills, and introducing students to AI concepts early. This requires educators and institutions to distill trends and skills into actionable knowledge that can be integrated into curricula. Due to the fast pace of AI developments and supporting technologies, frequent updates to education and training materials are essential. This ensures that students and professionals are prepared for the ongoing evolution.

The Role of Academic Institutions in Workforce Preparation

Colleges, universities and academic institutions are responsible for shaping the future workforce. By conducting research and integrating AI into their curricula, they ensure that graduates possess the skills required by the wider industry. Collaboration with industry partners allows academic programs to align with the evolving demands of the job market, creating a seamless transition for students entering the workforce.

The Role of the Semiconductor Education Alliance

Recognizing the scale of the challenge, initiatives like the Semiconductor Education Alliance (SEA) have emerged to bridge the gap between academia, industry, and government agencies. SEA's work focuses on developing knowledge, skills, and abilities (KSA) frameworks, which help by providing rigorous descriptions of the competencies required for software and hardware engineers. These frameworks are crucial for designing high-quality curricula and training materials that align with industry needs. The global community of SEA members also emphasize the importance of the "3 Cs" – content, community, and careers – to help build a robust ecosystem that prepares individuals for AI-driven roles.

Arm Collaborations with Academia

Arm has been actively working to catalyze industry-academia collaborations. By developing free educational tools and training kits, Arm has facilitated the widespread adoption of its content across thousands of universities. Meanwhile through SEA, Arm continues to work closely with academic institutions, contributing to educational standards and linking education with industry requirements.

More recently, Arm announced critical investment into the University of Cambridge's new CASCADE (Computer Architecture and Semiconductor Design) Centre, which will fund 15 Ph.D. students over the next five years who will undertake groundbreaking work in intent-based programming to realize the potential of AI through next-generation processor designs. Such academic collaborations ensure a steady pipeline of talent and innovations equipped to meet the demands of an AI-enabled world.

A Comprehensive Approach to Equipping the AI Workforce

This ecosystem approach complements the organizational culture strategies outlined in the main chapter. While Hinkle focuses on building AI readiness through internal training, change management, and inclusive approaches within organizations, these efforts are most effective when supported by the broader educational infrastructure discussed here. The integration of AI into the workforce requires a concerted effort to address the current skills gap. By revising curricula, investing in training, and fostering collaboration between academia, industry, and government, we can prepare the future workforce for success. Initiatives like SEA, and industry-leaders like Arm show how the collaborative spirit needed to adapt to this transformative era. This ensures that the global workforce is equipped to thrive in understanding, developing and using AI applications.

CHAPTER 8

Case studies

Streamlining ADAS Integration for Safer, Smarter Vehicles with LeddarTech

INTRODUCTION

As the automotive industry undergoes rapid transformation, vehicle OEMs face mounting pressure to accelerate development cycles and integrate advanced safety technologies like advanced driver-assistance aystems (ADAS). In the U.S. alone, ADAS is estimated to prevent up to 62% of total traffic deaths annually.

ADAS technologies rely on various sensing systems, such as cameras, LiDAR, radar, and ultrasonic sensors, combined with AI-driven algorithms to help detect potential hazards, assist drivers, and even take corrective action autonomously. These systems are essential for creating safer roads and reducing human error behind the wheel.

THE CHALLENGE

Despite the proven benefits of ADAS, OEMs and Tier 1 suppliers face several barriers to seamless integration. ADAS platforms must be compatible with a wide range of electronic control units (ECUs) and configurations from legacy vehicle development, requiring scalable solutions that work across both high-end and mass-market vehicles. Moreover, integrating ADAS into existing vehicle architectures demands complex software coordination capable of processing vast sensor inputs in real time.

In addition to hardware and software complexities, OEMs must comply with different global standards and regulations for ADAS implementation. Each region imposes varying requirements for compliance, making it challenging to ensure consistent safety and performance worldwide.

THE SOLUTION

To simplify ADAS adoption and improve efficiency, LeddarTech and Arm collaborated to optimize LeddarVision™, an AI-driven sensor fusion and perception software stack. This solution enables vehicles to create an accurate 3D environmental model using advanced AI and computer vision algorithms.

LeddarVision is built on Arm-based automotive platforms, which offer a standardized ADAS software framework that can be adapted across multiple platforms without extensive customization. This unified architecture integrates Arm advanced CPUs and accelerators to deliver optimized performance, while minimizing computational bottlenecks to result in improved system efficiency.

One of the key performance-enhancing features of LeddarVision is the use of Arm Cortex-A720AE CPUs, which provide a significant boost in

processing power for critical sensor fusion algorithms. This improvement reduces perception latency, enhancing the responsiveness of ADAS features and overall vehicle safety.

By leveraging the Arm pre-silicon port in the Armv9 architecture, LeddarTech is able to prototype and validate software enhancements before hardware deployment. This approach helps accelerate development cycles and ensure compatibility with future hardware advances, providing OEMs with an accelerated path to market.

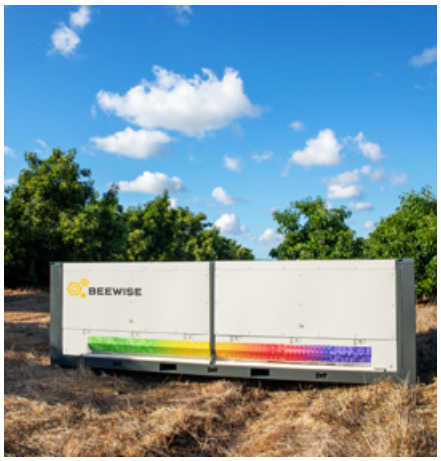
THE RESULT

LeddarVision, running on Arm-powered next-generation ECUs, delivers a range of benefits to OEMs and Tier 1 suppliers. The optimized software stack helps reduce CPU utilization, minimizing the need for high-power systems-on-chip (SoCs) while ensuring reliable ADAS performance. This improved computational efficiency leads to accelerated reaction times and reduced end-to-end latency, resulting in safer, more responsive ADAS functionality.

The standardized ADAS architecture simplifies the integration process across different vehicle models, helping to reduce complexity and lower development costs. In addition, LeddarVision undergoes rigorous real-world testing to ensure that its AI-driven perception models perform reliably in dynamic driving conditions.

By optimizing LeddarVision with Arm-based platforms, LeddarTech and Arm are setting a new standard for intelligent mobility. This collaboration helps to enable accelerated ADAS adoption, improved system efficiency, and enhanced vehicle safety, shaping the future of advanced automotive technology.

The First Autonomous Beehive: How Beewise is Revolutionizing Beekeeping with AI



INTRODUCTION

Bee populations are declining at an alarming rate, with 35% of bees dying annually due to climate change, pesticide exposure, and disease. Since bees pollinate one-third of the global food supply, their decline poses a major risk to agriculture and biodiversity.

Traditional beekeeping relies on manual inspections and wooden hives, making it slow, labor-intensive, and ineffective against modern threats. Beewise has developed Beehomes, the world's first AI-powered, robotic beehive that automates hive management and helps reduce bee mortality rates by 85%.

THE CHALLENGE

For centuries, beekeepers have used the same manual methods to manage their hives. They physically inspect colonies to check for diseases, pests, and environmental damage, but these visits happen only every few weeks. By the time an issue is detected, a colony may already be collapsing.

One of the biggest threats to honeybees is the Varroa mite, a deadly parasite that spreads viruses and weakens entire hives. Without early detection and treatment, an infestation can wipe out a colony within

weeks. Additionally, climate change has led to more extreme weather conditions, including rising temperatures and sudden cold snaps, further destabilizing bee populations.

Beekeepers also face economic and labor challenges. The manual nature of beekeeping is time-consuming, requiring significant resources and expertise. Honey harvesting disrupts bee colonies, affecting productivity and overall hive stability. A scalable, automated solution is needed to protect bees, reduce human intervention, and optimize hive management.

THE SOLUTION

Beewise's Beehomes are AI-driven, solar-powered hives that monitor, regulate, and protect bee colonies 24/7. Unlike traditional wooden hives, Beehomes use robotics, machine learning, and computer vision to provide real-time hive insights and automate critical processes.

These intelligent hives track temperature, humidity, and colony activity, automatically adjusting conditions for hive stability. The system detects early signs of disease and removes Varroa mites before infestations escalate. By using precision robotics, Beehomes automate honey harvesting, allowing for minimal disruption to the bees while maximizing production.

At the core of Beehomes is an Arm-based CPU, which enables real-time automation and AI-driven decision-making. NVIDIA Jetson processors analyze hive data, while Raspberry Pi modules provide cloud connectivity, allowing beekeepers to monitor their hives remotely. Solar-powered operation helps ensure that the system remains completely self-sufficient, reducing reliance on external energy sources.

Beehomes replace guesswork with data-driven precision, giving

beekeepers an advanced tool to manage their colonies effectively, while requiring far less manual intervention.

THE RESULT

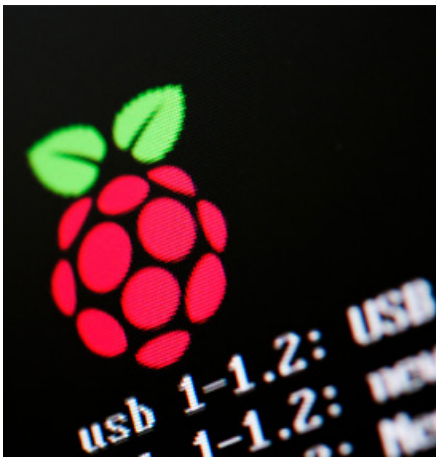
Since Beehomes were introduced, beekeepers have seen an 85% reduction in bee mortality rates, a game-changing improvement for colony survival. Healthier bees have led to higher honey yields, increasing productivity while maintaining hive stability.

By eliminating unnecessary manual interventions, Beewise has cut labor costs and made beekeeping more scalable and efficient. The ability to detect and address threats in real time has prevented colony collapses that would have been unavoidable under traditional methods.

Beehomes are also proving to be a sustainable solution for large-scale beekeeping operations. With their solar-powered design and energy-efficient architecture, they align with eco-friendly practices, while helping to ensure long-term hive stability. Beekeepers can now manage multiple hives remotely, freeing up time to focus on expansion rather than crisis management.

By merging AI, robotics, and automation, Beewise is helping to secure the future of beekeeping and global food production. As pollinators continue to face threats, Beewise innovation helps ensure that bees can thrive. Saving bees means saving ecosystems, agriculture, and the planet.

Accessible Computing Platform For Everyone: The Evolution of Raspberry Pi



INTRODUCTION

Fifteen years ago, the Raspberry Pi Foundation was established with a mission to democratize computing for young people. Founded by Eben Upton, the initiative aimed to create an affordable and accessible computing platform to empower students and makers worldwide. What began as an educational project soon evolved into a powerful computing tool, extending far beyond classrooms into industrial and embedded applications.

Today, Raspberry Pi is a critical platform in the industrial and embedded electronics sector, serving a diverse range of users across industries.

THE CHALLENGE

In the past, the computing industry struggled with accessibility and affordability. Traditional hardware solutions were expensive, proprietary, and primarily designed for large-scale enterprises, leaving students, hobbyists, and small businesses without viable options. The absence of an affordable, flexible computing platform stifled innovation, restricting opportunities for learning, experimentation, and custom development. At the same time, industries seeking embedded computing solutions needed compact, power-efficient, and scalable alternatives to costly proprietary systems. The market lacked a truly accessible and adaptable computing solution

that could serve as a versatile computing platform for education, hobbyists, and industry alike.

THE SOLUTION

Raspberry Pi, a compact, flexible, and cost-effective computing platform, emerged as the solution to this accessibility gap. Originally developed to support education, it quickly found adoption in industrial and embedded applications. Today, companies worldwide integrate Raspberry Pi into a variety of applications, including factory automation, smart manufacturing, medical devices, IoT systems, and autonomous monitoring solutions, leveraging its compact form factor, power efficiency, and adaptability to diverse computing needs.

A major enabler of Raspberry Pi's success is its partnership with Arm, which has played a massive role in delivering performance, energy efficiency, and scalability. From the Arm11 family of processors in the original Raspberry Pi to the Arm Cortex-A76, a high-performance processor optimized for modern computing tasks in Raspberry Pi 5, this collaboration has enabled Raspberry Pi to continuously advance in power and capability. The integration of Arm technology has ensured continuous innovation, making Raspberry Pi an essential platform for educators, developers, and businesses alike.

"Arm's technology allows us to offer cutting-edge AI capabilities in a way that's accessible to everyone." –Eben Upton, CEO, Raspberry Pi Foundation

Beyond education, companies worldwide integrate Raspberry Pi hardware into embedded products, industrial systems, and IoT solutions, leveraging its power, flexibility, and affordability. This synergy between Raspberry Pi and Arm continues to push the boundaries of accessible computing.

THE RESULT

Raspberry Pi has completely revolutionized computing accessibility, becoming a key player in industrial automation, AI development, and real-world problem-solving. Developers and engineers worldwide rely on Raspberry Pi to build autonomous systems, AI-driven analytics, edge computing frameworks, and connected devices. Its affordability has empowered researchers, small businesses, and students to experiment with robotics, environmental monitoring, and smart home solutions, proving that cost-effective computing can be powerful and scalable.

The partnership with Arm continues to drive Raspberry Pi's innovation, enabling it to deliver high-performance, low-power computing across industries. From supporting machine learning and AI deployments to fueling next-generation industrial control systems, Raspberry Pi remains a pioneer in advancing digital transformation. With its commitment to providing flexible, powerful computing platforms, the Raspberry Pi Foundation is shaping the future of embedded computing, helping to ensure technology remains accessible to all.

SpaceTech Smart City Infrastructure Powered by Arm's Edge AI

INTRODUCTION

The future of urban living depends on intelligent, data-driven solutions that help enhance safety, sustainability, and efficiency in real time.

SpaceTech, a subsidiary of China Vanke Co., Ltd., is a leading smart city solutions provider that manages more than 8 million residential units and 2,000 commercial buildings, integrating technologies to optimize urban living spaces. To streamline operations and improve the resident experience, SpaceTech has integrated Arm-based computing solutions into its infrastructure, enabling real-time decision making, automated building management, and energy efficiency enhancements.

THE CHALLENGE

With a vast portfolio that spans over 1 billion square meters, SpaceTech faced several key challenges in managing smart urban environments.

One of the major concerns was waste management, where inefficient manual oversight led to delays in service. AI-driven categorization and automated service requests were required to streamline operations. Another challenge lay in legacy building systems, where disparate devices, such as lighting, HVAC, and access control, operated in isolation. The lack

of interoperability between devices from different manufacturers created inefficiencies, making centralized management difficult.

Additionally, the growing volume of data traffic from smart devices demanded a high-performance, power-efficient solution to process AI workloads at the edge, while keeping energy consumption low.

THE SOLUTION

To address these challenges, SpaceTech partnered with Ampere Computing to transition from x86 edge servers to Arm-based Ampere Altra servers. This move provided significant advantages, including 2.5 times improved performance per rack, 2.8 times lower power consumption, and a 3 times smaller footprint, aligning with SpaceTech's sustainability goals.

A proof of concept demonstrated the efficiency of the Alibaba Qwen-VL Vision Language Model (VLM) on Ampere Arm-based servers, showing significant energy savings without compromising AI capabilities. Ampere also provided dedicated AI libraries with optimized quantization methods, improving performance by up to 2 times, while maintaining model accuracy.

In smart waste management, AI-driven categorization now automatically detects full bins and generates service tickets.

In building automation, real-time AI analytics integrate various systems, allowing lighting, HVAC, and security to function as a unified network. Motion sensors detect room occupancy, adjusting energy use dynamically, while surveillance AI enhances safety by monitoring elevators, preventing unauthorized scooter use, and managing cleanliness in public areas.

THE RESULT

The transition to Arm-powered edge AI servers has enabled SpaceTech to create a scalable, high-efficiency smart city infrastructure. The implementation of intelligent automation has led to a significant reduction in energy consumption, while improving security and operational efficiency.

With Arm-based edge computing, SpaceTech has streamlined multiple building systems, running various applications in virtual containers on edge server clusters. This unified data-lake approach has eliminated inefficiencies, allowing seamless integration across lighting, HVAC, air quality, and access control systems.

Additionally, the edge cloud-native platform developed by SpaceTech helps ensure that every device—from cameras to sensors to gateways—functions as a micro-cloud, enabling real-time AI decision making, while reducing power and cooling demands. The adoption of Arm-powered solutions positions SpaceTech to lead the smart city revolution, setting new benchmarks for intelligent urban management.

By embracing Arm technology and Ampere Arm-based edge servers, SpaceTech is redefining urban property management with scalability, AI-powered efficiency, and sustainability, offering a model for the future of connected smart cities worldwide.

CONCLUSION

From AI Readiness to AI Leadership

As we've seen throughout this report, AI is reshaping business operations globally at an unprecedented pace. With 82% of business leaders already using AI applications and 87% planning to increase their AI budgets over the next three years, the momentum is undeniable. Yet a critical readiness gap persists—only 39% have a clearly defined AI strategy, and just 29% can automatically scale computing resources to meet AI demands.

The Arm AI Readiness Index reveals that organizations face challenges across multiple dimensions. From infrastructure limitations (only 23% have dedicated power infrastructure for AI workloads), to talent shortages (34% report they are under-resourced with AI talent), to data readiness issues (46% cite data quality as a major barrier), these challenges must be addressed systematically.

Security concerns add another layer of complexity, with 48% of leaders worried about data privacy breaches through model extraction. As personally identifiable information increasingly fuels AI systems,

organizations must implement robust security frameworks while addressing ethical considerations like bias detection and correction.

The chapters in this report provide a roadmap for organizations seeking to thrive in an AI-powered future. We've explored the technical foundations required for AI success, the evolving regulatory landscape, approaches to managing AI safety and risk, security frameworks that build trust, sustainability considerations, and strategies for developing an AI-ready culture.

Organizations that will lead in the AI era aren't merely adopting technology—they're transforming their entire approach to business, building AI ecosystems that balance innovation with environmental responsibility. They're implementing governance structures that ensure compliance while fostering ethical AI development. They are creating cultures where employees at all levels understand AI's potential and contribute to its responsible implementation.

The difference between AI followers and AI leaders will increasingly be defined not by the technology itself, but by how organizations prepare their infrastructure, develop their talent, manage their data, and address emerging risks. By taking a comprehensive approach to AI readiness—one that spans technology, people, processes, and governance—organizations can transform AI from a promising technology into a powerful engine for sustainable innovation and competitive advantage. The future belongs to those who prepare for it today.

The AI Readiness Index shows where we stand now. The question for every organization is: Where will you be tomorrow?

Appendix

References

01 AI Risk Management Framework. NIST [Internet]. 2021 Jul 12 [cited 2025 Feb 20]; Available from: https://www.nist.gov/itl/ai-risk-management-framework

02 AI Verify Foundation [Internet]. [cited 2025 Feb 20]. What is AI Verify. Available from: https://aiverifyfoundation.sg/what-is-ai-verify/

03 AI Watch: Global regulatory tracker – Brazil | White & Case LLP [Internet]. 2024 [cited 2025 Feb 20]. Available from: https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-brazil

04 Artificial Intelligence and Data Act (AIDA) – Companion document [Internet]. Innovation, Science and Economic Development Canada; 2025 [cited 2025 Feb 20]. Available from: https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document

05 Arm. A leap forward in auto safety with LeddarTech and Arm [Internet]. Available from: https://armkeil.blob.core.windows.net/developer/Files/pdf/case-study/a-leap-forward-in-auto-safety-with-leddartech-and-arm.pdf

06 Arm. First autonomous beehive: Beewise uses AI to protect pollinators [Internet]. Available from: https://www.arm.com/company/success-library/made-possible/first-autonomous-beehive
Arm. Raspberry Pi: From education tool to AI-enabled edge computing [Internet]. Available from: https://www.arm.com/company/success-library/made-possible/raspberry-pi

07 Arm Edge AI technology powers SpaceTech's next-generation smart cities. Arm Blog [Internet]. 2025 Jan 17. Available from: https://newsroom.arm.com/blog/arm-edge-ai-powers-spacetech-smart-cities

08 Arm Newsroom Blog. AWS Graviton Decarbonize Compute. Arm Blog [Internet]. 2023 November 21. Available from: https://newsroom.arm.com/blog/aws-graviton-decarbonize-compute

09 Arab News/AFP. The software behind the self-driving Uber crash did not recognize jaywalkers [Internet]. 2019. Available from: https://www.arabnews.com/node/1579826/science-technology

10 BBC News. Google partners with Kairos Power to deploy small nuclear reactors for AI data centers [Internet]. 2024 Sep. Available from: https://www.bbc.com/news/articles/c748gn94k95o

11 Bellamy RKE, Dey K, Hind M, Hoffman SC, Houde S, Kannan K, et al. AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. IBM J Res Dev. 2019;63(4/5):4:1–4:15. doi: 10.1147/JRD.2019.2942287

12 Biggio B, Corona I, Nelson B, Rubinstein BIP, Maiorca D, Fumera G, et al. Evasion attacks against machine learning at test time. Machine Learning and Knowledge Discovery in Databases. 2013;8190:387–402.

13 Biggio B, Nelson B, Laskov P. Poisoning attacks against support vector machines. In: Proceedings of the 29th International Conference on Machine Learning (ICML). 2012;1807–14.

14 Bradford A. The Brussels Effect: How the European Union Rules the World. Oxford University Press; 2020.

15 Brookings Institution. Explainability won't save AI [Internet]. 2020. Available from: https://www.brookings.edu/articles/explainability-wont-save-ai

16 Brookings Institution. Global networked AI governance [Internet]. (No date). Available from: https://www.brookings.edu/articles/network-architecture-for-global-ai-policy/

17 California Consumer Privacy Act (CCPA). California Civ. Code § 1798.100 et seq.

18 Campbell G. Arm and industry leaders launch Semiconductor Education Alliance to address the skills shortage [Internet]. Arm Newsroom; 2023 Aug 1. Available from: https://newsroom.arm.com/news/semiconductor-education-alliance

19 China. Data Security Law. 2021.

20 China. Ethical Guidelines for AI. 2022.

21 China. Personal Information Protection Law. 2021.

22 China Privacy Law | Office of Ethics, Risk and Compliance Services [Internet]. Available from: https://oercs.berkeley.edu/privacy/international-privacy-laws/china-privacy-law

23 China State Council. New Generation Artificial Intelligence Development Plan. 2017.
'New Generation Artificial Intelligence Development Plan' (2017)" in List 1)

24 CIODive. California governor vetoed SB 1047. What's next for AI regulation? [Internet]. 2024 Sept. 30. Available from: https://www.ciodive.com/news/california-senate-bill-SB1047-AI-regulation-landscape/728481

25 Clark L. GenAI's dirty secret: It's set to create a mountainous increase in e-waste. The Register. 2024 Oct 28. Available from: https://www.theregister.com/2024/10/28/genai_dirty_secret/

26 CNN Business. Microsoft to restart reactor at Three Mile Island for AI power by 2028 [Internet]. 2024 Sep 20. Available from: https://edition.cnn.com/2024/09/20/energy/three-mile-island-microsoft-ai/index.html

27 Cornell University (INFO 2040 course blog). The 2010 Flash Crash: information cascades [Internet]. 2020. Available from: https://blogs.cornell.edu/info2040/2020/11/13/the-2010-flash-crash-how-information-cascades-shape-our-world/

28 Cyberspace Administration of China. Internet Information Service Algorithmic Regulation Provisions. 2022.

29 Dastin J, Nellis S. Focus: For tech giants, AI like Bing and Bard poses billion-dollar search problem [Internet]. Reuters; 2023 Feb 23. Available from: https://www.reuters.com/technology/tech-giants-ai-like-bing-bard-poses-billion-dollar-search-problem-2023-02-22

30 Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT). 2019;4171–4186.

31 Deloitte. Powering artificial intelligence: A study of AI's environmental footprint—today and tomorrow. November 2024. Available from: https://www.deloitte.com/global/en/issues/climate/powering-ai.html

32 Dung L. Current Cases of AI Misalignment and Their Implications for Future Risks. Synthese. 2023;202:138. doi:10.1007/s11229-023-04367-0

33 European Commission. EU AI Act – Regulatory framework [Internet]. 2024. Available from: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

34 European Commission. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance) [Internet]. Jun 13, 2024.

35 European Union. AI Act. Official Journal of the European Union; August 2024.

36 Falade PV. Decoding the threat landscape: ChatGPT, FraudGPT, and WormGPT in social engineering attacks. Int J Sci Res Comput Sci Eng Inf Technol. 2023;8(5):185–200. doi: 10.32628/CSEIT2390533

37 Garcia-Tobin C, Knight M. Elevating security with Arm CCA. Commun ACM [Internet]. 2024 Oct;67(10):34–39. Available from: https://cacm.acm.org/practice/elevating-security-with-arm-cca

38 Gallup. Strategy Will Fail Without a Culture That Supports It [Internet]. 2024. Available from: https://www.gallup.com/workplace/652727/strategy-fail-without-culture-supports.aspx

39 Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Nets. In: Advances in Neural Information Processing Systems (NeurIPS). 2014;27:2672–2680.

40 Government of Canada. Bill C-27: The Artificial Intelligence and Data Act. 2022.

41 Hadshar R. A Review of the Evidence for Existential Risk from AI via Misaligned Power-Seeking [Preprint]. arXiv; 2023 Oct 27. Available from: https://arxiv.org/abs/2310.18244

42 He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV; 2016. pp. 770–778. doi: 10.1109/CVPR.2016.90

43 Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Comput. 1997;9(8):1735–1780. doi: 10.1162/neco.1997.9.8.1735

44 ICO (UK). Google DeepMind and NHS patient data [Internet]. ICO. Available from: https://ico.org.uk/for-the-public/ico-40/google-deepmind-and-class-action-lawsuit/

45 Illinois Biometric Information Privacy Act (BIPA). 740 Ill. Comp. Stat. 14/1 et seq.

46 Infocomm Media Development Authority (Singapore). Model AI Governance Framework. 2019; updated 2020.

47 International Energy Agency. Data centres and data transmission networks [Internet]. Available from: https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks

48 ISO & IEC. ISO/IEC 42001: AI Management Systems – Requirements. 2023.

49 K8sGPT. AI-Powered resource and cluster orchestration [Internet]. 2024. Available from: https://k8sgpt.ai/

50 Kabashkin I. End-to-end service availability in heterogeneous multi-tier cloud-fog-edge networks. Future Internet. 2023;15:329.

51 Kilian KA. Beyond Accidents and Misuse: Decoding the Structural Risk Dynamics of Artificial Intelligence [Preprint]. arXiv; 2024 Jun 21. Available from: https://arxiv.org/abs/2406.14873

52 KPMG. ISO/IEC 42001 AI Management Systems [Internet]. 2023. Available from: https://kpmg.com/ch/en/insights/artificial-intelligence/iso-iec-42001.html

53 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS'12). Curran Associates Inc.; 2012. pp. 1097–1105.

54 Laux J, Wachter S, Mittelstadt B. Trustworthy artificial intelligence and the European Union AI Act: On the conflation of trustworthiness and acceptability of risk. Regul Gov. 2024;18(1):3–32. doi: 10.1111/rego.12512.

55 Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems. 2017;30:4765–4774.

56 Mayer Brown (Bloomberg Law). Conducting an AI Risk Assessment [Internet]. 2024. Available from: https://www.mayerbrown.com/en/insights/publications/2024/01/conducting-an-ai-risk-assessment

57 McMahan HB, Moore E, Ramage D, Hampson S, Arcas BA. Communication-efficient learning of deep networks from decentralized data. arXiv [Preprint]. 2017 Feb 17. Available from: https://arxiv.org/abs/1602.05629

58 Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. ACM Comput Surv. 2021;54(6):Article 115. doi: 10.1145/3457607

59 Microsoft. Tay chatbot incident [Internet]. 2016. Available from: https://en.wikipedia.org/wiki/Tay_(chatbot)

60 Millière R. The Alignment Problem in Context [Preprint]. arXiv; 2023 Nov 3. Available from: https://arxiv.org/abs/2311.02147

61 Nemko. Ensuring AI Safety and Robustness [Internet]. (No date). Available from: https://www.nemko.com/blog/ai-safety-and-robustness

62 New America OTI. Global cooperation – International Network of AI Safety Institutes [Internet]. 2024. Available from: https://www.newamerica.org/oti/blog/to-achieve-global-ai-cooperation-we-need-multidisciplinary-dialogues-in-addition-to-the-international-network-of-ai-safety-institutes

63 Personal Data Protection Commission (Singapore). AI Verify Toolkit. 2022.

64 Radford A, Narasimhan K, Salimans T, Sutskever I. Improving Language Understanding by Generative Pre-Training. OpenAI; 2018.

65 Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, et al. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. J Mach Learn Res. 2020;21(140):1–67.

66 Raffin E, Hamidouche W, Nogues E, Pelcat M, Menard D, Tomperi S. Energy efficiency of a parallel HEVC software decoder for embedded devices. In: Proceedings of the 12th ACM International Conference on Computing Frontiers (CF '15). New York: Association for Computing Machinery; 2015. pp. 1–6. doi: 10.1145/2742854.2747286

67 Reuters. Amazon scrapes secret AI recruiting tools that show bias against women [Internet]. 2018. Available from: https://www.ml.cmu.edu/news/news-archive/2016-2020/2018/october/amazon-scraps-secret-artificial-intelligence-recruiting-engine-that-showed-biases-against-women.html

68 Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD Int'l Conference on Knowledge Discovery and Data Mining. 2016:1135–1144.

69 Silver D, Schrittwieser K, Simonyan K, Antonoglou I, Huang A, Guez A, et al. Mastering the game of Go without human knowledge. Nature. 2017;550(7676):354–359. doi: 10.1038/nature24270

70 Singh RK, Mishra S. TinyML meets IoBT against sensor hacking. In: Proceedings of the 2024 Workshop on Security and Privacy in Standardized IoT. 2024. doi: 10.14722/sdiotsec.2024.23010

71 Sofia RC, et al. A framework for cognitive, decentralized container orchestration. IEEE Access. 2024;12:79978–80008. doi: 10.1109/ACCESS.2024.3406861

72 Securiti.ai. Overview of Australia's AI Safety Standard [Internet]. 2023. Available from: https://securiti.ai/australia-voluntary-ai-safety-standard/

73 Srivatsa M, Abdelzaher T, He T. Artificial intelligence for edge computing. Springer; 2023. 365 p. Available from: https://doi.org/10.1007/978-3-031-40787-1

74 TechMonitor. Frontier Model Forum launched for safe AI [Internet]. 2023. (OpenAI, Google, Microsoft, Anthropic partnership).

75 TIME. Exclusive: New Research Shows AI Strategically Lying [Internet]. 2024 Dec 18. Available from: https://time.com/7202784/ai-research-strategic-lying

76 Time.com. AI's growing demand for data-center electricity [Internet]. 2024. Available from: https://time.com/6987773/ai-data-centers-energy-usage-climate-change/

77 UNFCCC. Paris Agreement. Bonn, Germany: UNFCCC; 2015. Available from: https://unfccc.int/sites/default/files/english_paris_agreement.pdf

78 University of Cambridge. Arm donates £3.5 million for Cambridge PhD students to study computer architecture and semiconductor design [Internet]. 2024. Available from: https://www.cam.ac.uk/news/arm-donates-ps3-5-million-for-cambridge-phd-students-to-study-computer-architecture-and

79 Voigt P, Von dem Bussche A. The EU General Data Protection Regulation (GDPR): A Practical Guide. Springer International Publishing; 2017.

80 Wikipedia. AI Alignment [Internet]. (No date). Available from: https://en.wikipedia.org/wiki/AI_alignment

81 World Economic Forum. Why upskilling your existing workforce can be far more cost-effective than hiring new employees [Internet]. 2024. Available from: https://www.weforum.org/stories/2024/01/ai-training-workforce/

82 Zhou I, Tofigh F, Piccardi M, Abolhasan M, Franklin DR, Lipman J. Secure multi-party computation for machine learning: A survey. IEEE Access. 2024;12:53881–53899.

83 Zhu J-Y, Park T, Isola P, Efros AA. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy; 2017. pp. 2242–2251. doi: 10.1109/ICCV.2017.244

84 Translation: Internet Information Service Algorithmic Recommendation Management Provisions – Effective March 1, 2022 [Internet]. DigiChina. [cited 2025 Feb 20]. Available from: https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022

85 Translate CL. Data Security Law of the PRC [Internet]. China Law Translate. 2021 [cited 2025 Feb 20]. Available from: https://www.chinalawtranslate.com/datasecuritylaw/

86 Position Paper of the People's Republic of China on Strengthening Ethical Governance of Artificial Intelligence (AI) [Internet]. [cited 2025 Feb 20]. Available from: https://www.fmprc.gov.cn/eng/zy/wjzc/202405/t20240531_11367525.html

87 Model AI Governance Framework 2024 - Press Release | IMDA [Internet]. [cited 2025 Feb 20]. Available from: https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2024/public-consult-model-ai-governance-framework-genai

Survey methodology

METHOD

Arm conducted this research using an online survey prepared by Method Research and distributed by RepData and iTracks among n=655 adults (age 18+) who are business leaders across various industries. All respondents are from companies with more than 1,000 employees and influence purchase decision-making within their organizations. The sample was from the following countries: U.S., U.K., Germany, France, Hungary, China, Taiwan and Japan. Data was collected from January 16 to February 5, 2025.

Research led by Svetlana Gershman, Vice President and the head of Method Communications' specialized marketing research division, and Fatima Aslam, Research Analyst at Method Communications.

SAMPLE DEMOGRAPHIC

COUNTRY

US - 15%

UK - 15%

China - 15%

France - 8%

Germany - 8%

Japan - 16%

Taiwan - 15%

AGE

Gen Z (18–27) - 13%

Millennial (28-43) - 43 %

Gen X (44-59) - 35%

Boomer (60-78) - 10%

GENDER

Man - 60%

Woman - 40%

DEPARTMENT

Finance -13%

Marketing/Sales - 14%

Accounting- 10%

Operations- 12%

Human Resources- 14%

Information Technology- 24%

Legal services - 13%

FUNCTIONAL ROLE

C-level executive- 17%

Vice President - 13%

Director - 36%

Manager - 34%

COMPANY SIZE

SMB - 44%

Mid Market - 28%

Enterprise - 28%

COMPANY ANNUAL REVENUE (US)

Less than $1 Million - 9%

$1 million - under $40 Million - 20%

$40 million - $1 billion - 37%

More than $1 billion - 24%

Prefer not to say - 12%

INDUSTRY

Financial Services - 18%

Manufacturing - 19%

Healthcare - 4%

Retail - 9%

Technology - 17%

Energy - 4%

Other - 29%

About the Contributors

EXTERNAL AUTHORS

Deval Shah

Senior Machine Learning Engineer at the Australian Institute for Machine Learning (AIML)

Deval Shah is a Senior Machine Learning Engineer at the Australian Institute for Machine Learning (AIML), where he specializes in designing Retrieval Augmented Generation (RAG) systems for enterprise search and policymaking applications. His focus on scalability and secure data handling underscores his expertise in production-ready AI solutions. Prior to AIML, Deval advanced from Machine Learning Engineer to Senior Software Engineer at Uncanny Vision, spearheading projects that enhanced license plate recognition models and vehicle re-identification microservices. Deval has collaborated with over 15 AI companies, published more than 100 articles, and successfully built substantial organic inbound channels for AI startups through his content-driven initiatives. His skill set spans full-stack development (Next. js, FastAPI), containerization (Docker), and cloud infrastructure (AWS, Azure DevOps).

Dr John Soldatos

Honorary Research Fellow at the University of Glasgow

John Soldatos holds a Ph.D. in Electrical & Computer Engineering from the National Technical University of Athens (2000) and is currently an Honorary Research Fellow at the University of Glasgow, UK (2014-present). He was

Associate Professor and Head of the Internet of Things (IoT) Group at the Athens Information Technology (AIT), Greece (2006–2019), and Adjunct Professor at the Carnegie Mellon University, Pittsburgh, PA (2007–2010). He has significant experience working closely with large multi-national industries (e.g., IBM, INTRACOM, INTRASOFT International) as an R&D consultant and delivery specialist while being a scientific advisor to various high-tech startup enterprises. Dr. Soldatos is an expert in Internet-of-Things (IoT) and Artificial Intelligence (AI) technologies and applications, including IoT/AI applications in smart cities, finance (Finance 4.0), and industry (Industry 4.0).

Mark Hinkle

CEO and Co-Founder Peripety Labs

Mark Hinkle has over 30 years of experience in technology, open source, and enterprise software. He served in executive roles at Cloud.com, Citrix and The Linux Foundation, spearheading innovative programs in cloud computing, DevOps, and emerging technology adoption. Over the past two years, he has trained thousands of professionals on how to use AI for real-world impact—ranging from beginners to C-suite executives. Mark publishes TheAIE.net, a newsletter network with more than 250,000 subscribers dedicated to working smarter with AI. He is also the co-founder of All Things AI, an organization focused on advancing enterprise AI through events, community, and industry collaboration.

Dr Nicole Höher

Project Manager for Sustainability & Digitalisation at juS.TECH GmbH

Dr Höher is a natural scientist with many years of experience in data analysis and visualisation. She specialises in the application of data science and artificial intelligence, and has experience in the development of AI

software solutions. Her expertise combines technical know-how with an understanding of environmental, economic and social contexts to develop innovative and sustainable solutions.

Dr Nora von Ingersleben Seip

Postdoctoral Researcher at the Political Science Department of the University of Amsterdam, RegulAite project

Nora von Ingersleben-Seip is a Postdoctoral Researcher at the University of Amsterdam, where she investigates international cooperation on AI governance. Previously, she was the Managing Director of a venture capital-funded tech startup in Southeast Asia and held other executive roles in the region's startup ecosystem. She also led political advocacy efforts for major global tech firms in North America and Europe. Her research on AI governance has been published in leading peer-reviewed journals and cited by The Wall Street Journal and Politico, among others. She is a member of several expert groups, including the AI Adoption Initiative and the Forum for Cooperation on Artificial Intelligence at the Brookings Institution and the Center for European Policy Studies. Nora holds a PhD (with Distinction) from the Technical University of Munich and an MBA from INSEAD.

Dr Vanessa Just

Founder and CEO of juS.TECH GmbH

Dr Just is an entrepreneur and expert in business, industrial engineering and artificial intelligence. As a leading voice on sustainability, digitalisation and AI, she speaks on numerous panels and events and is a lecturer at the FOM University of Applied Sciences. Dr Just is also an author and editor for renowned publishers such as Springer: "Digitalisation and Sustainability" and "Sustainability through Innovation, Digitalisation and Technologies". Dr Just is also a board member of the German AI Association.

Method Communications

Research for the AI Readiness Index survey and narrative for Chapter 1 led by Svetlana Gershman, Vice President and the head of Method Communications' specialized marketing research division, and Fatima Aslam, Research Analyst at Method Communications.

ARM AUTHORS

Kevork Kechichian

Executive Vice President, Solutions Engineering

Kevork is leading the team that works closely with Arm's world-class partner ecosystem to develop and accelerate our technologies as we keep ahead of changing market needs.

His career spans 30 years in various semiconductor engineering and leadership roles. Before joining Arm, he was the executive vice president of MCU/MPU Engineering at NXP Semiconductors, where he led the company's global architecture and IP & SoC design. Previously, he was the senior vice president of engineering at Qualcomm where he managed the Snapdragon SoC and Technology teams and was responsible for delivering more than 100 SoCs in leading-edge technologies, enabling breakthrough products in mobile and consumer markets. Kevork holds a BEng degree in electrical engineering from American University of Beirut and a M.S. in electrical engineering from Concordia University.

Khaled Benkrid

Senior Director, Education and Research

With a background that spans both academia and industry, Benkrid has been instrumental in addressing the skills gap in the semiconductor

industry. In his current position, he leads Arm's Education and Academic Engagements Team, leveraging his experience to bridge the divide between academic education and industry needs. He also holds the position of Visiting Professor in Computing at Anglia Ruskin University (ARU) in Cambridge, England.

Benkrid has been actively involved in several key initiatives, including playing a formative role in the creation of the Semiconductor Education Alliance and working with organizations such as the Semiconductor Research Corporation (SRC) in the U.S., the Taiwan Semiconductor Research Institute (TSRI), and the All India Council for Technical Education (AICTE) to develop necessary semiconductor workforce, especially in light of recent initiatives like the US CHIPS Act.

Benkrid is also a prolific writer and thought leader in the field of education and technology. He has authored articles on topics such as "The Evolving Nature of Work" and "Talent Multiplication: The Holy Grail of Modern Organizations," discussing the impact of AI on the job market and the importance of continuous learning. Benkrid holds a Ph.D. and an MBA and is a Chartered Engineer (CEng).

Maureen McDonagh

Head of Sustainability

Maureen is passionate about driving Arm's sustainability journey in this transformational era of AI, balancing the opportunities and challenges for Arm to contribute to a sustainable, equitable and resilient future. Energized by Arm's incredible talent of global problem solvers and the role we can all play at the intersection of technology, innovation, people, and planet. Connecting ideas, people, resources to catalyse the new thinking we need to solve our environmental and social challenges to shape a more sustainable and inclusive world.

Vince Jesaitis

Senior Director, Government Affairs

Vince is a seasoned government affairs professional with over 20 years of experience in the high-tech sector. He currently serves as the Senior Director of Government Affairs at Arm, the world's leading semiconductor and software design company.

Vince joined Arm in October 2021 and is responsible for developing and executing advocacy campaigns, building relationships with key policymakers and influencers, and representing the company's vision and values in various forums.

Prior to his role at Arm, Vince held senior positions at the Information Technology Industry Council and in the U.S. Congress, where he worked on several legislative initiatives benefiting the high-tech sector. Vince holds a Bachelor of Arts degree in International Relations, Philosophy, and Political Science from William Jewell College, which he earned between 1997 and 2001.

Will Abbey

Executive Vice President and Chief Commercial Officer

Will leads sales, field engineering, and partner enablement at Arm, helping innovative organizations harness the latest technologies to prepare for the next wave of AI-driven digital transformation. From IP to AI, his insight has guided technology leaders in transforming their products and operations.

Since joining Arm in 2004, Will has held several leadership roles, including General Manager of the Physical Design group. Will has held product management roles at Celoxica, Infineon Technologies and Loughborough Sound Images, and serves on the board of EnPro Industries.

Additional Arm Contributors

Arm Content

Brian Fuller

Jack Melling

Omkar Patwardhan

Kurt Wilson

Arm Branding

Julie Jervis

Sarah Nguyen

Arm Creative

Chris Salvador

Hoa Tong

Arm Sustainability Team

Mohammed Chunara

Fiona Maudslay

Fiona Riggall

Wevolver Editorial team

Editor: Rebecka Durén

Operations: Jessica Miley

We would like to extend our sincere gratitude for the dedication and hard work of everyone who contributed to the creation of this report. Your expertise, collaboration, and commitment made this possible—thank you!

About Arm

Arm is the industry's highest-performing and most power-efficient compute platform with unmatched scale that touches 100 percent of the connected global population. To meet the insatiable demand for compute, Arm delivers advanced solutions that allow the world's leading technology companies to unleash the unprecedented experiences and capabilities of AI. Together with the world's largest computing ecosystem and 20 million software developers, we are building the future of AI on Arm.

www.arm.com

arm

About Wevolver

Wevolver is a global platform and community that provides engineers with the knowledge and connections to develop better technology.

We bring a professional audience of engineers informative and inspiring content, such as articles, videos, podcasts, and reports, about state-of-the-art technologies.

The knowledge on Wevolver is published by various sources: universities, tech companies, and individual community members. Next to that, we manage a network of over 50 technical writers who create content for our customers and publish that on Wevolver.com

Millions of engineers leverage Wevolver to stay up to date, find knowledge when they are developing products, and leverage the platform to make meaningful connections.

Wevolver has won the SXSW Innovation Award, the Accenture Innovation Award, and the Top Most Innovative Web Platforms by Fast Company. Wevolver is how today's engineers stay cutting edge.

www.wevolver.com