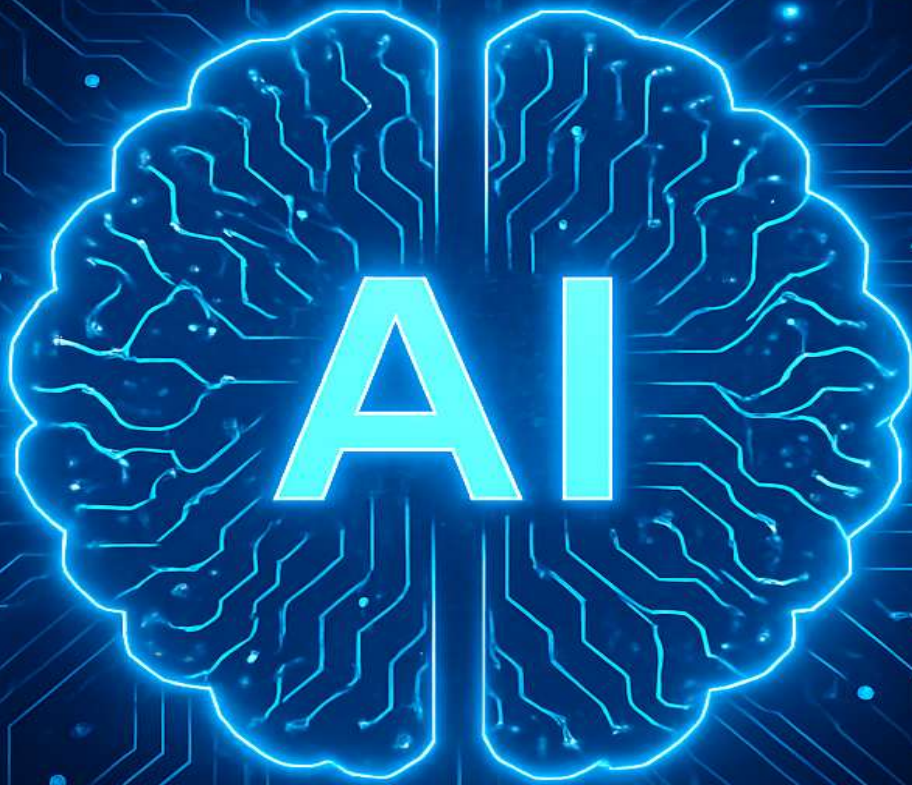


Credible AI Demands Credible Data
January 2025



This report is strictly confidential and has been prepared based on independent analysis conducted by Crownleaf specialists, complemented by insights from local consultants with country-specific expertise. Crownleaf is a privately held research and advisory firm operating independently, with no institutional or political affiliation to governments, political parties, multilateral organizations, or regulatory bodies. The opinions expressed in this document reflect the technical interpretation of the data available at the time of writing and are subject to change as new information emerges. The content herein does not constitute investment advice, legal counsel, or political guidance, and is intended solely for the internal use of authorized clients and strategic partners. No part of this report may be reproduced, distributed, or published, in whole or in part, without the express written consent of Crownleaf. Unauthorized use may lead to legal liability. Crownleaf Inc. © 2025

Introduction

Artificial intelligence (AI) has emerged as a top global priority. Efforts to establish shared principles and safeguards for AI now unfold against a backdrop of deepening geopolitical rivalry, policy fragmentation, and limited international coordination.

Yet, despite these tensions, progress has been notable. Initiatives such as the UN Global Digital Compact and the Paris AI Action Summit illustrate how governments are recognizing the shared responsibility to guide AI's evolution responsibly. The challenge ahead lies in aligning and integrating the growing constellation of frameworks already shaped by organizations like the OECD, UNESCO, the Council of Europe, and the G7—each contributing valuable perspectives, but still lacking unified coherence.

At the heart of trustworthy AI lies the issue of data governance—a critical yet often siloed dimension of policymaking. While AI development depends fundamentally on high-quality, interoperable data, discussions around global data flows and governance have struggled to match the momentum of broader AI discourse. Worse, some countries are introducing data localization and sovereignty policies that risk fragmenting global data ecosystems and stifling AI innovation.

To fully realize AI's transformative potential, a more integrated approach is needed—one that connects AI governance, data policy, and technological infrastructure. That means:

- Ensuring data is accurate, secure, and globally accessible;
- Revitalizing multilateral efforts that support open yet responsible data exchange;
- Creating policy environments that encourage innovation without overregulation.

As this report outlines, resolving foundational data issues is essential to building trustworthy AI. It explores emerging business strategies, policy considerations, and governance pathways that can unlock growth while navigating the geopolitical complexities of the digital age.

Why Data Quality Is Essential for Artificial Intelligence

Data is the essential fuel powering today's AI revolution. Alongside rapid advances in algorithms and hardware, access to massive datasets has enabled powerful technologies like machine learning to flourish. The sheer volume of available data is staggering—with projections estimating over 221 zettabytes by 2026, enough to store 110 trillion¹ movies.

Training AI models relies on nearly all publicly available internet content, alongside proprietary datasets from industries and governments. The indexed web alone holds approximately 500 trillion words, expected to grow by 50% by 2030². Yet, even as data volumes surge, they may not keep pace with the growing demands of larger and more complex AI models.

Developers are now exploring richer data sources:

- Multimodal datasets combining text, video, and images;
- Synthetic data³ generated by AI itself.

At current trajectories, data limitations—not chips or energy—may become the dominant constraint on AI innovation by the end of the decade. In this context, high-quality data becomes not just beneficial, but indispensable.

Beyond Big Data: Why Quality Matters

For specialized applications and smaller language models (SLMs), data quantity is far less valuable than its precision and relevance. These use cases require curated, trustworthy data with robust governance. Ironically, while generative AI expands capabilities, it also contributes to deteriorating data quality online—blurring the line between authentic and AI-generated content.

Poor data can lead to serious consequences:

- Hallucinations and inaccurate outputs;
- Model drift and reduced reliability;
- Costly financial misjudgments, like in the case of a major U.S. real estate platform that incurred over \$500 million in losses due to flawed pricing algorithms. Their heavy reliance on historical data, without factoring local differences, amplified the errors across thousands of transactions.

These risks are magnified in sensitive sectors:

- Healthcare;
- Transportation;
- Financial services.

In these areas, precision is non-negotiable—yet many commercially available models aren't optimized for such high-stakes environments. Organizations increasingly adopt multimodel approaches to ensure robustness and relevance.

¹ [csis.org/analysis/operationalizing-data-free-flow-trust-dfft](https://www.csis.org/analysis/operationalizing-data-free-flow-trust-dfft)
² epoch.ai/blog/can-ai-scaling-continue-through-2030
³ arxiv.org/html/2211.04325v2

To address the growing risks associated with AI deployment, many companies have embraced responsible AI principles and practices. These standards are not static—they're continuously reviewed and updated to align with evolving global regulations and best practices.

These companies approach reflects key international governance frameworks, including:

- The EU AI Act;
- The OECD AI Principles and AI definition;
- The AI Risk Management Framework (RMF) from the U.S. National Institute of Standards and Technology (NIST).

These frameworks emphasize the need for accurate, transparent, and auditable data—especially critical in the financial services sector, where trust and precision are the foundation of successful operations.

For financial institutions, the integrity of AI systems hinges on the reliability of the data they use. Ensuring that data meets high standards is not just about regulatory compliance—it's about safeguarding reputations, making sound decisions, and maintaining market confidence.

be reluctant to raise their own climate commitments. As of now, many countries have yet to submit their nationally determined contributions (NDCs) for 2035. And among those that are expected, there's little indication that the upcoming pledges will demonstrate substantially greater ambition or provide the policy clarity needed to encourage emerging markets to act with confidence

The Economic Power of Quality Data

Artificial intelligence is emerging as a general-purpose technology—on par with game-changers like electricity and the internet. Its adaptability across virtually every industrial sector gives it enormous potential to boost productivity, ignite economic growth, and tackle complex, real-world challenges.

One key enabler of AI's impact is specialized datasets, which allow organizations to fine-tune foundation models for targeted applications. For instance, agricultural innovator John Deere leverages AI models trained on decades' worth of geospatial and soil quality data to differentiate crops from weeds in real time. This precision enables smart spraying techniques, cutting costs for farmers and consumers alike, while reducing environmental harm—a prime example of AI driving operational efficiency and sustainability simultaneously.

Industry-specific and proprietary datasets are empowering traditional businesses to harness the potential of AI, resist disruption, and uncover new revenue streams. This shift enables companies—large and small, and far beyond the tech sector—to tap into the tangible benefits of AI, rather than leaving the rewards concentrated in Silicon Valley.

Many organizations, especially those with deep operational histories, have accumulated high-quality data over decades. These datasets are rich in context, unique in scope, and irreplaceable by synthetic alternatives. When integrated with AI technologies, they unlock new possibilities for customer engagement, product development, and decision-making.

As AI technology matures, frontier models are beginning to converge in capability and become widely available at increasingly lower costs. Analysts suggest that this trend will lead to the commoditization of models, making them more interchangeable and less of a competitive differentiator.

In this evolving landscape, unique, high-quality datasets will emerge as the true engine of value creation. Rather than relying solely on off-the-shelf large language models (LLMs), businesses will integrate AI more deeply across operations—ushering in a wave of AI diffusion that enhances productivity, drives innovation, and transforms industry sectors.

The strategic importance of proprietary data lies in its ability to:

- Enable fine-tuned, domain-specific AI applications;
- Support differentiated services and insights;
- Fuel long-term economic growth beyond model advancements.

As AI adoption spreads across the global economy, data access and quality will become critical levers of competitiveness—not just for businesses, but also for nations. Countries with rich, well-governed datasets will gain an edge in innovation, improve self-sufficiency, and position themselves strongly in the global AI race.