InfoBrief, sponsored by AWS | September 2023

Putting Responsible AI into Practice

BEST PRACTICES AND GUIDELINES

Ritu Jyoti
Group Vice President, Worldwide Artificial Intelligence and Automation Research Practice, Global AI Research Lead, IDC

Table of Contents

CLICK BELOW TO NAVIGATE TO EACH SECTION IN THIS DOCUMENT.

Executive Summary

Artificial intelligence (AI) and digital transformation (DX) have become one of IT's dynamic duos, transforming businesses worldwide. IDC predicts that direct DX investments will accelerate to a compound annual growth rate (CAGR) of 16.5% for 2022 to 2024, up from a CAGR of 15.4% for 2019 to 2024, making up 55% of all information and communication technology investment by the end of 2024. **AI spending is expected to grow to $301 billion at a CAGR of 26.5% for 2021 to 2026.**

As awareness grows regarding potential risks associated with deploying AI (traditional and generative), including legal, cyber, privacy, intellectual property, toxicity, hallucinations, ethical, or cultural concerns, building and **deploying responsible AI has become a priority for organizations across all sectors.** With the right guardrails, filters and responsible AI principles, organizations will be able to enjoy the business benefits and outcomes of generative AI.

AI Investments Are at the Forefront of Enterprise Digital Transformation

AI and machine learning (ML) are everywhere. Adding generative AI accelerates innovation for startups, creating entirely new business domains and business models.

of survey respondents are investing in AI/ML solutions to future-proof their business.

- Reduce staffing/labor costs
- Improve sustainability
- Reduce production and delivery costs
- Get better insights
- Improve customer and employee experiences
- Increase employee productivity
  - Decisions
  - Innovation

n = 2,053, Base = worldwide; Source: IDC's *AI StrategiesView 2022*, May 2022

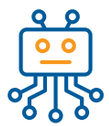
AI Investments Are at the Forefront of Enterprise Digital Transformation (continued)

Augmented customer service

- Provides 24 x 7 customer support and self-service in a natural way
- Empowers human agents with contextual guidance

IT optimization

- Orchestrates the linking of IT systems to become self-acting and self-regulating
- Automates mundane software maintenance activities



AutoML

- Automates the tasks of applying machine learning to real-world problems
- Includes every stage from beginning with a raw data set to building a model ready for deployment

Supply and logistics

- Optimizes and augments digital supply chain operations, providing end-to-end visibility
- Predicts and optimizes the flow of goods to track raw material, completed products
- Predicts supply chain risk from planning and execution system data, improving supply chain resiliency

Sales next best action

- Supports sales professionals with AI-optimized next best actions based on sales stage, win probability, and best practices learned from other opportunities

Fraud analysis and investigation

- Helps detect illegal or illicit financial acts involving intentional deception or misrepresentation across different areas (i.e., operational and financial) of an organization

n = 2,053, Base = worldwide; Source: IDC's *AI StrategiesView 2022*, May 2022

AI Can Give Rise to Unwanted, Sometimes Serious Concerns

Responsible AI focuses on reducing unintended consequences of AI by aligning the system's intent and use with the norms and values of the users it aims to serve.

As the reach and adoption of AI/ML continue to grow, responsible AI is increasingly top of mind for business decision makers, policy makers, data scientists, and business analysts alike. The responsible creation and use of AI solutions is an important business imperative to build trust.

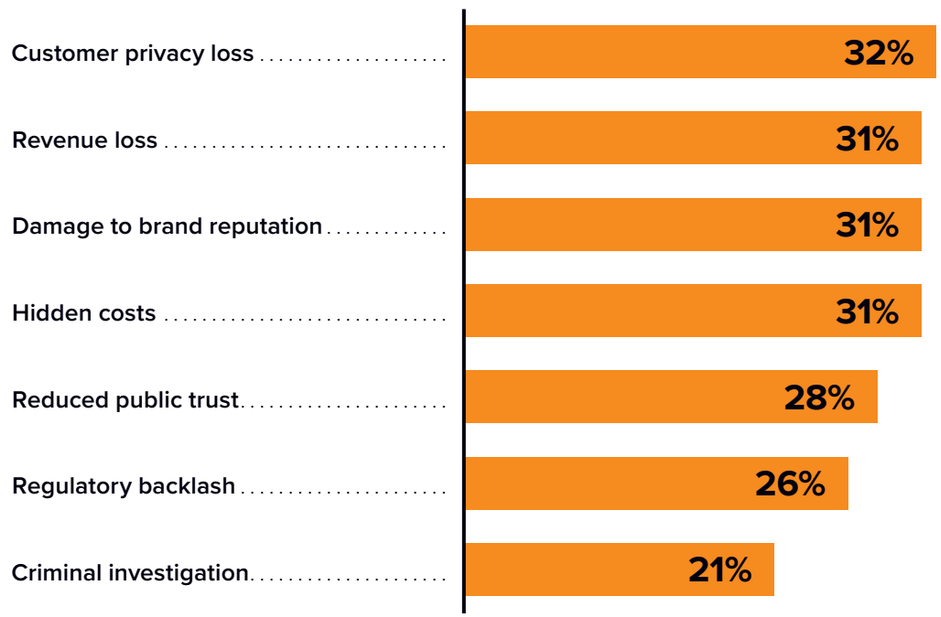
AI/ML Business Risk

- Data privacy
- Bias
- Black box
- Noncompliance
- Safety

Potential Negative Business Impact

- Damage to brand reputation
- Reduced public trust
- Revenue loss
- Regulatory backlash
- Customer privacy loss
- Hidden cost

What are your top two potential negative business impacts if the AI/ML is not implemented responsibly?



n = 2,017; Source: IDC's *AI StrategiesView 2022*, May 2022

Companies Struggle to Operationalize Responsible AI and AI Governance

Fairness — How do we make sure AI systems perform equally well across groups?

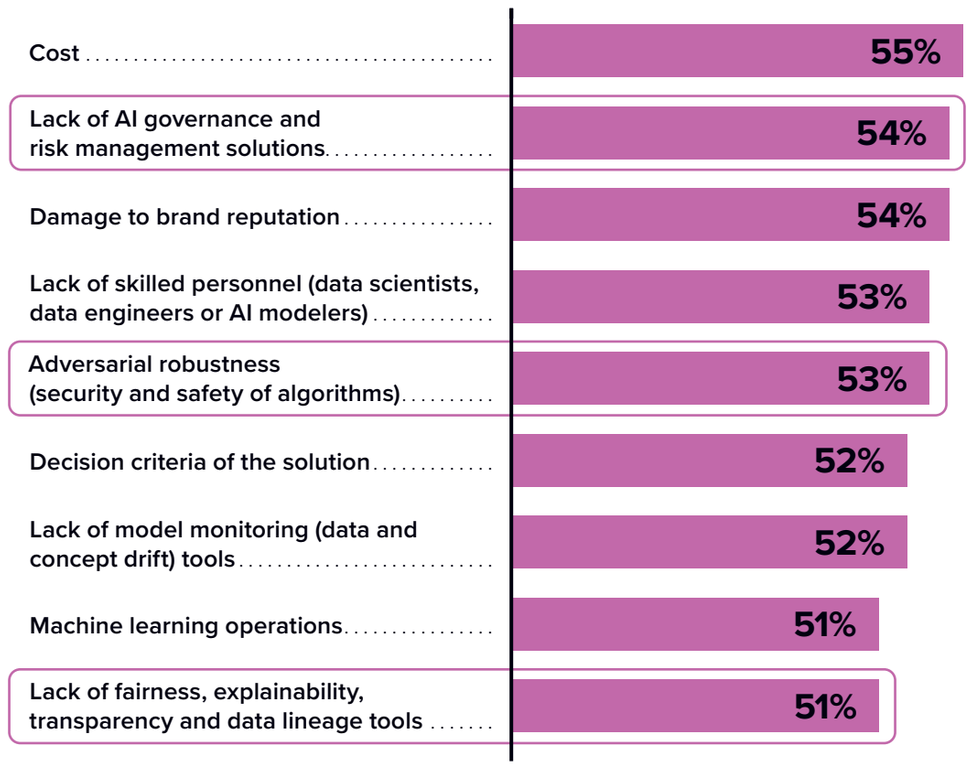
Explainability — If AI systems are opaque and unable to explain how or why certain results are presented, this lack of understanding will undermine trust in the system.

Adversarial robustness — AI systems should be safe and secure, ensure privacy, and not be vulnerable to tampering or compromising the data they are trained on.

Lineage — AI systems should include details of their development (e.g., data usage), deployment, and maintenance so they can be audited throughout their life cycle.

Transparency — Disclosures (reporting in action) and transparency (fact sheets) in AI systems are nascent areas of research but are key to the mainstream adoption of AI.

What are the main challenges for implementing AI technology at your organization?



n = 2,053, Base = worldwide; Source: IDC's *AI StrategiesView 2022*, May 2022

Emerging Challenges with Generative AI

Businesses need to be mindful of new issues as they explore and embrace use cases.

Increased potential for misuse, biases, and toxicity

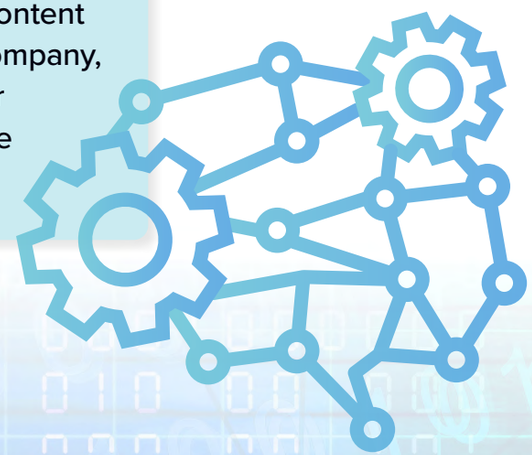
Like traditional AI, generative AI can display bias, largely due to bias in its data. Yet the risk can be more intense, since generative AI may also create misinformation and abusive or offensive content.

Black box

Generative AI usually runs on a "foundation model" built by a specialized third party. Since the business does not own this model or have access to its inner workings, understanding why it produced a particular output may be impossible. Organizations must test foundation models adequately and mitigate any unintended consequences.

Cybersecurity threats

While generative AI can help produce compelling content, it can help malicious actors to do the same. For example, generative AI could be used to produce content that appears to be from a company, spreading misinformation or urging stakeholders to share sensitive data.



Emerging Challenges with Generative AI (continued)

Novel data security and privacy challenges

Generative AI's ability to connect data in its vast data sets (and to generate new data as needed) could compromise privacy controls without proper protections. By identifying relationships among apparently disparate data points, generative AI could identify stakeholders who have been made anonymous and piece together their sensitive information.

Hallucinations that impact performance

Generative AI is good at presenting authoritative answers to almost any question. Yet sometimes its answers are flat-out wrong, which data scientists call "hallucinations." Hallucinations occur in part because the models are designed to generate content that seems reasonable; that doesn't mean the output will be accurate.

Intellectual property and copyright complexities

With so much data underlying generative AI, it's not always possible to know its source—or if there is permission to use it. Generative AI may reproduce copyrighted text, images, or software code in the content that it produces in the user's name.

Inadvertent sharing of proprietary data

A company's proprietary data and insights could help competitors generate content. For example, as organizations enter their proprietary data to fine-tune a generative model, they should consider whether they want that data to then be used to train the model more broadly.

Regulations Governing AI Vary Across Countries

United States

- A fragmented approach to AI regulation has resulted in states enacting a patchwork of laws.
- The U.S. Congress enacted the **National AI Initiative Act** in 2021, creating "an overarching framework to strengthen and coordinate AI research, development, demonstration, and education activities across all U.S. Departments and Agencies."
- The act created new offices and task forces aimed at implementing a national AI strategy involving a multitude of U.S. administrative agencies, including the Federal Trade Commission, Department of Defense, Department of Agriculture, Department of Education, and the Department of Health and Human Services.
- Pending legislation includes the **Algorithmic Accountability Act of 2022,** which was introduced in both houses of Congress in February 2022. In response to reports that AI systems can lead to biased and discriminatory outcomes, the proposed act would direct the Federal Trade Commission to create regulations that mandate "covered entities," including businesses meeting certain criteria, to perform impact assessments when using automated decision-making processes.

European Union

- In the EU, the European Commission has published an overarching **regulatory framework proposal** titled the Artificial Intelligence Act.
- Applications are sorted into risk categories: minimal, limited, high, or unacceptable. An application's risk level determines government action or obligations, such as enhancing the security, transparency, and accountability of AI applications through human oversight and ongoing monitoring.
- Companies will be required to register standalone high-risk AI systems, such as remote biometric identification systems, in an EU database.
- If the proposal is passed, the earliest date for compliance would be the second half of 2024. Potential fines range from 2%–6% of a company's annual revenue.
- The previously enacted EU General Data Protection Regulation (GDPR) already carries implications for AI technology. **Article 22** prohibits decisions based on solely automated processes that produce legal consequences or similar effects for individuals unless the program gains the user's explicit consent or meets other requirements.

Asia/Pacific

- In March 2022, China passed a regulation governing companies' use of algorithms in online recommendation systems, requiring that such services be moral, ethical, accountable, transparent, and "disseminate positive energy."
- Companies must notify users when an AI algorithm is playing a role in determining which information to display to them, and users must have the option to opt out of being targeted.
- The regulation prohibits algorithms that use personal data to offer different prices to consumers.
- The Monetary Authority of Singapore suggests that firms should establish approval for highly material AI decisions at the CEO or board level. The company should periodically update the board on its use of AI, enabling the board to maintain a central view of all material AI-driven decisions.

Laws regulating AI are rapidly evolving. Global businesses must comply with the differing standards that are emerging from APAC, the European Union, and the United States. It is important to note that these jurisdictions are not comprehensive. Regulators in Hong Kong and the Netherlands have been outspoken on the need for appropriate corporate governance to address AI-related risks, including those related to bias, model drift, privacy, cybersecurity and transparency, and operational failures.

Businesses Need AI Governance Solutions with Multiple AI Capabilities

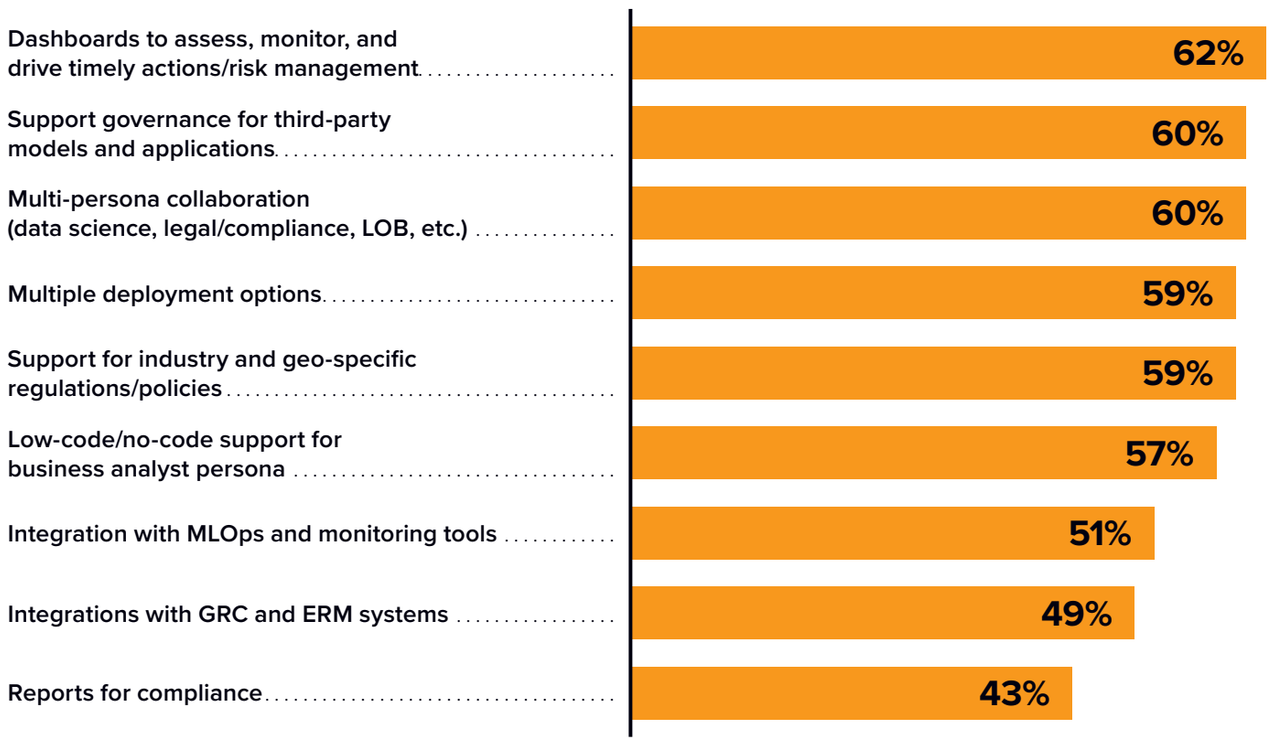
Foundations of AI Governance

1. AI/ML life-cycle governance

2. Risk management

3. Regulatory excellence

What are the critical capabilities of an AI governance solution?



n = 2,053, Base = worldwide; Source: IDC's *AI StrategiesView 2022*, May 2022

Evolve from a Reactive Approach to a Proactive One

Responsible AI Governance Starts at the Executive Level

The C-suite has critical roles and responsibilities.

Chief Financial Officer

- AI cost and financial risk

Chief Marketing Officer

- AI for customer and brand charters

Chief Data Officer

- The evolution of AI governance charter and data governance

Chief Legal and Compliance/Risk Officer

- AI legal and risk factors for the organization

Chief Executive Officer

- The AI governance charter and organizational accountability

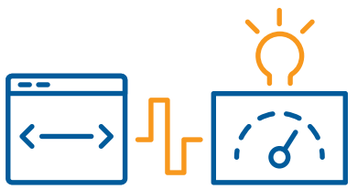
Chief Human Resources Officer

- The creation of AI employee policy and charter

Poor AI governance increases the risk of unintended negative consequences but is complicated by endlessly changing regulations.

Foundations of AI Governance

1. Ensure AI/ML Life-Cycle Governance



Data scientists oversee the machine learning operations.

Model validators and approvers are responsible for:

- Monitoring the model in production for fairness, drift, quality, and explainability
- Automated collection of metadata about model development and experiments, including training data
- Model validation/approval metadata collection

2. Incorporate Collaborative Risk Management

Risk management is an integral part of corporate governance, risk, and compliance (GRC). Companies must ensure that AI models adhere to AI regulations and that a policy engine can define and enforce automated AI governance policies and rules.

Framework for AI Risk Management

DEFINE
Set up risk management GRC standards

IMPLEMENT
Apply GRC standards to all AI projects

VERIFY
All AI projects follow various rules and policies

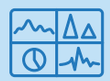
OPERATE
Repeat monitoring and certification every time model changes

Fairness, explainability, and robustness > Human oversight > Societal and environmental well-being > Internal and external accountability

3. Strive for Regulatory Excellence for AI

- Model testing and validation
- Data preparation
- Exploration
- Model training and experimentation

External data science/ development

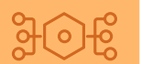


In-house data science/ development

- Production deployment
- Model life-cycle management
- Analytics and reporting
- Business integrations

IT, MLOps, DevOps, Security

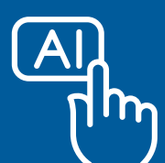
Model risk management



- Model certification
- Risk management
- Regulations and impact assessment
- Model audit and reporting

External auditors, FDA, others

Internal compliance/ audit



Regulatory excellence for AI comes through digital collaboration across an organization and by exploiting leading-edge tools and technologies.

Why AWS for Responsible AI

AWS and Responsible AI

From theory to practice

AWS provides the tools, guidance, and resources customers need to get started and implement responsible AI across their organization.

- The "Responsible Use of Machine Learning Guide" provides considerations and recommendations for responsibly developing and using ML systems across three major phases of their life cycles.
- AI Service Cards are a new resource to increase transparency and help customers better understand AWS AI Services.
- AWS offers purpose-built capabilities: Amazon SageMaker Clarify, Amazon SageMaker Model Monitor, ML Governance tools, Amazon Augmented AI, and more.

Educate the next generation of leaders

AWS is educating the next generation of ML leaders to help promote fairness and mitigate bias.

- AWS AI & ML Scholarship program helps underserved and underrepresented high school and college students learn foundational ML concepts to prepare for careers in AI and ML.
- AWS Machine Learning University offers a new bias mitigation and fairness course with over nine hours of lectures and hands-on exercises available free.
- Strategic partnerships with the University of California, Berkeley; MIT; the California Institute of Technology; the University of Washington; and others.

The science of responsible AI

AWS is working to advance the science behind responsible AI with active research and development.

- Amazon has invested $20 million in the National Science Foundation Fairness in AI Grants program.
- Deep engagement with multi-stakeholder organizations including the Global Partnership on AI, Responsible AI Institute, The Partnership on AI, and National Artificial Intelligence Advisory Committee to share best practices, accelerate research, and responsibly develop AI and ML technology.
- Ongoing research grants are made through the Amazon Research Awards and scientific publications with Amazon Science.

A Holistic Approach to Building AI Services Responsibly

AWS integrates responsible AI across the end-to-end machine learning life cycle.

Iterative, continuous process

At AWS, AI is an integral part of the entire machine learning life cycle, including design and development, deployment, and ongoing use.

Responsible AI is an iterative process that requires continuous testing and auditing for potential bias and accuracy.

Expert teams

AWS has teams of dedicated experts wholly committed to staying on the cutting edge of research, providing best practices, and developing rigorous methodology to build AWS AI and ML services in a responsible way.

AWS engineering and development teams audit the algorithms powering AI services and routinely conduct internal testing to develop products that are accurate and fair.

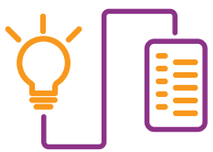
Secure, private data

Customers benefit from AWS datacenters and a network architected to protect their information, identities, and applications. Customers always own their data and are able to encrypt it, move it, and manage retention.

AWS works to ensure that customer data remains private and secure, with good governance to control who has access.

AWS Responsible AI Tools and Capabilities

AWS offers innovative tools and capabilities that customers can leverage at all stages of the AI/ML life cycle to help build, train, and operate systems responsibly.



Detect bias and explain model predictions

Amazon SageMaker Clarify detects potential bias during data preparation, after model training, and in deployed models. SageMaker Clarify also provides greater visibility into model behavior so customers can provide transparency to stakeholders and track whether a model is performing as intended.

Enable monitoring and human review

Amazon SageMaker Model Monitor automatically detects and alerts you to inaccurate predictions from models deployed in production, and with Amazon Augmented AI, you can implement human review of ML predictions when human oversight is needed.

Improve governance

SageMaker Role Manager, SageMaker Model Cards, and SageMaker Model Dashboard improve governance of ML projects with tighter control and deeper visibility over ML models. You can set up users with least-privilege permissions, easily capture and share model information, and stay informed on model behavior, such as bias, all in one place.

Enhance transparency

AWS AI Service Cards are a resource to help customers better understand AWS AI services and use them responsibly. They provide information on the intended use cases and limitations, responsible AI design choices, and deployment and performance optimization best practices. The first three AI Service Cards are:

- Amazon Rekognition—Face Matching
- Amazon Textract—Analyze ID
- Amazon Transcribe—Batch (English-US)

Building Generative AI in a Responsible Way

AWS builds foundation models (FMs) with responsible AI in mind at each stage of its comprehensive development process.

Steps AWS takes to build responsibly:

- **Introduce guardrails.** Amazon's Titan FMs are built to detect and remove harmful content in the data that customers provide for customization, reject inappropriate content in the user input, and filter the model's outputs containing inappropriate content (such as hate speech, profanity, and violence).
- **Keep data secure and private.** Customer data is not used to train the original base models. All data is encrypted, and customers can configure their Amazon Virtual Private Cloud settings to access Amazon Bedrock APIs and provide model fine-tuning data in a secure manner. In this way, customers can trust that their data will remain private and confidential.

Amazon CodeWhisperer is the only AI coding companion with built-in security scanning (powered by automated reasoning) for finding and suggesting remediations for hard-to-detect vulnerabilities.

Generative AI technology will continue to evolve, posing new challenges that will require additional attention and mitigation with collaboration across academic, industry, and government partners.

Customer Success Scenarios

AUTOMATIC DATA PROCESSING

ADP is a comprehensive global provider of cloud-based human capital management solutions that unite HR, payroll, talent, time, tax and benefits administration, as well as outsourcing services and compliance.

The company was struggling to scale its adoption of enterprise AI applications and had concerns about potential bias, transparency, and projects delivering erroneous outcomes due to bias.

ADP partnered with AWS Professional Services to not only get trusted insights about building good data sets that impact responsible AI, but also to operationalize bias monitoring and data privacy in tools and techniques. With Amazon SageMaker, ADP can deploy, monitor, and manage its machine learning pipelines.

Customer Success Scenarios

NATWEST GROUP

NatWest Group, a major financial services institution, partnered with AWS Professional Services to build a new MLOps platform and used Amazon SageMaker to create standardized end-to-end MLOps processes. This reduced the time to value for ML solutions from 12 months to less than three and reduced costs while maintaining high security and auditability.

A core part of NatWest's vision is to help customers thrive, which they can only do if decisions are made in a fair, equitable, and transparent manner. For ML models, this requires model explainability, bias reporting, and performance monitoring. The business can track and understand how and why models make specific decisions.

Due to lack of standardization across teams, no consistent model monitoring capabilities were in place, and teams had to create a lot of custom solutions. Two vital components of the new SageMaker project templates were bias detection on the training data and monitoring the trained model.

Amazon SageMaker Model Monitor and Amazon SageMaker Clarify met these requirements, operating in a scalable and repeatable manner.

Essential Guidance

We have now entered the world of AI-augmented work and decision making across all the functional areas of a business, from front to back office. AI, machine learning, and natural language processing are changing brands around the globe across multiple industry sectors. **AI disrupters will drive better customer engagements and have faster rates of innovation, higher competitiveness, higher margins, and superior employee experiences.** Organizations worldwide must evaluate their vision and transform their people, processes, technology, business models and data readiness to unleash the power of AI and thrive in the digital era. However, responsible AI implementation continues to be a major challenge.

An organization that wishes to accelerate the AI adoption and time to value with responsible AI should:

- Establish an organizationwide responsible AI strategy aligned with business goals.
- Be data driven with a focus on mitigating bias and improving data quality.
- Equip stakeholders for responsible use and oversight of generative AI. Teach employees the basics of how generative AI works, when and how to use it, and when and how to verify or modify outputs. Ensure compliance and legal teams keep up with the latest legal and regulatory developments and leverage the appropriate tools like regulatory policy packs along with GRC systems as applicable.
- Create organizational transparency with strong ongoing AI governance methodologies.
- Create a cross-functional group of AI experts to proactively address AI-specific risks and biases aggressively. This group should include IT as well as those in business and compliance functions.
- Coordinate the drivers for change for responsible AI inside and outside of the organization. Governance guidelines should clearly address why they are in place and leave little room for interpretation. Teams across different functions, such as leadership, data science, and legal, must understand the imperatives and incentives of each.
- Partner with a trusted and innovative technology supplier and professional services firm like AWS. AWS is focused on helping customers transform responsible AI from theory to practice. Look for ongoing innovations from AWS in responsible AI to help you reimagine your business.

About the IDC Analyst

Ritu Jyoti
Group Vice President
Worldwide Artificial Intelligence and Automation
Research Practice, Global AI Research Lead, IDC

Ritu Jyoti is group vice president, Worldwide Artificial Intelligence and Automation Research with IDC's software market research and advisory practice. She is responsible for leading the development of IDC's thought leadership for AI research and management of the Worldwide AI and Automation Software research team. Her research focuses on the state of enterprise AI efforts and global market trends for the rapidly evolving AI and machine learning innovations and ecosystem. Ritu also leads insightful research that addresses the needs of the AI technology vendors and provides actionable guidance on how to crisply articulate their value proposition, differentiate, and thrive in the digital era.

More about Ritu Jyoti

Message from the Sponsor

Transform responsible AI from theory into practice with AWS.

- Explore the new AWS AI Service Cards, a resource to increase transparency and help customers better understand our AWS AI Services:
  - Amazon Rekognition — Face Matching
  - Amazon Textract — AnalyzeID
  - Amazon Transcribe — Batch (English-US)
- Get started with purpose-built capabilities to help detect bias in data sets and models, better monitor and review model predictions, and improve governance:
  - Amazon SageMaker Clarify
  - Amazon SageMaker Model Monitor
  - ML governance with Amazon SageMaker
- Check out AWS Machine Learning University's new free hands-on course on bias mitigation and fairness, featuring over nine hours of lectures and exercises.

Learn more about AWS's approach to responsible AI

- Delve into the science behind the emerging challenges and solutions to build generative AI responsibly.

Read the Amazon Science blog post

IDC Custom Solutions

IDC Research, Inc.
140 Kendrick Street, Building B, Needham, MA 02494, USA
T +1 508 872 8200

@idc @idc idc.com

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. With more than 1,300 analysts worldwide, IDC offers global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries. IDC's analysis and insight helps IT professionals, business executives, and the investment community to make fact-based technology decisions and to achieve their key business objectives.