AI EXECUTIVE ROUNDTABLE

Deploying High-Quality and Trustworthy AI

Insights from Leading AI Practitioners

WHITEPAPER

Table of Contents

Introduction - Trustworthy AI in Practice 3

1. The state of AI adoption across industries 5

2. Attitude towards emerging regulations and certification schemes 7

3. Elements of a robust AI Quality framework 11

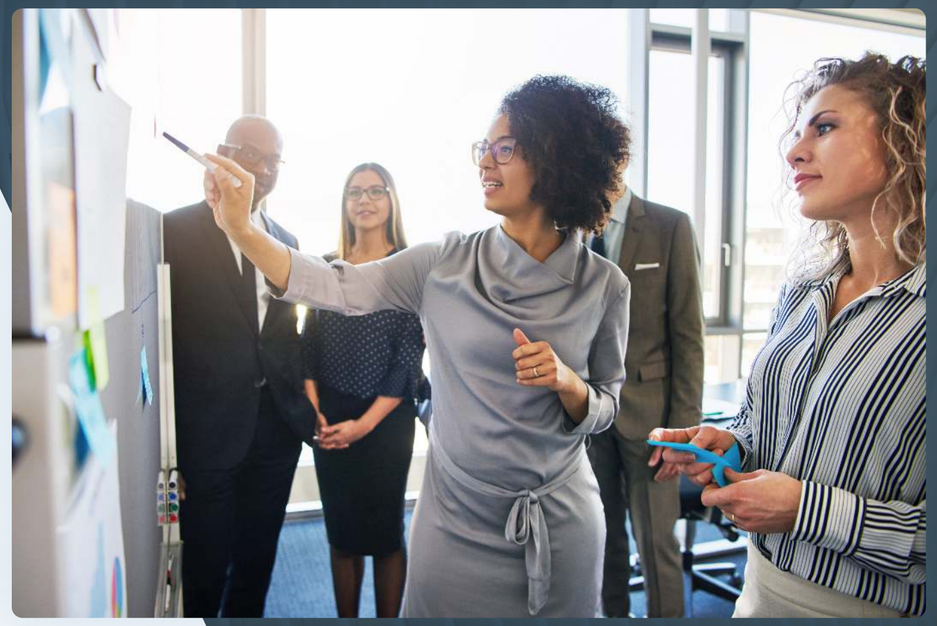
4. Steps to scale Trustworthy AI 14

Conclusion 16

Lead Author 17

Introduction

Trustworthy AI in Practice



AI has huge transformational potential in many industries, but most organisations have found it difficult to scale up adoption. One key barrier is the issue of the trustworthiness of applications based on artificial intelligence; or, more pointedly, the lack of trustworthiness of those applications. In 2021, the European Commission released the draft of its Artificial Intelligence Act – a comprehensive regulatory proposal that highlights the importance of transparency, reliability, robustness, fairness, and other characteristics of a high-quality AI system. Industry-level regulators, such as the **ECB**, and national regulators, like **Banque De France**, have made similar recommendations. The EU AI Act is expected to take effect in 2025, and forward-looking companies are already preparing for its implementation.

Much ground has been covered around the conceptual frameworks to deploy Trustworthy AI, but little is known about how industry actors are responding to this challenge. In practice, how do they embed AI quality and trustworthiness requirements into their model development, testing, and deployment processes? How do approaches vary by industries and geographies? How are leading AI adopters setting the pace?

Nozha Boujemaa
Global VP, Responsible AI
IKEA Retail

Danny Lange
Senior VP of AI
Unity

Joydeep Sarkar
Chief Analytics Officer
Holmusk

Yves Nicolas
Deputy Group CTO
Sopra Steria

Shameek Kundu
Head Of Financial Services and Chief Strategy Officer
TruEra

Practitioners' forums to share perspectives on these important questions are rare but essential to the industry to make progress on the trustworthy AI challenge. In September 2022, TruEra organized the first edition of the EMEA AI Executive Roundtable, an exclusive, invitation-only virtual event for data science, technology, and business executives who are driving AI adoption within their companies. The event also included representatives from international organizations and standard setting bodies. It featured a panel discussion (see panelists on the left inside) on the challenges in achieving real-world success from AI, followed by moderated, networking roundtables.

The primary goal was to facilitate peer-to-peer learning discussions between AI/ML industry executives across EMEA, and selected counterparts from the United States, on how to make AI initiatives successful in the real world, through a focus on AI quality and trustworthiness. This whitepaper captures the key insights from that event, spanning a plenary session and seven thematic roundtables, involving over 100 AI practitioners across Europe and beyond.

1. The state of AI adoption across industries

Investment into AI and ML continues at pace, across industries.

Across industries and geographies, participants confirmed that investment into the people, processes, and technology needed to build and deploy AI/ML systems remains a priority. Even in the context of highly-regulated industries, we heard of extensive use of AI/ML, ranging from back-office automation initiatives in financial services to real-world evidence generation in healthcare and pharma.

Technology-first vs. technology-enabled business models show clear differences in the scale and speed of AI adoption.

On one side are the technology-first "new" businesses – across sectors such as e-commerce, ride-hailing, search, and gaming – for which large-scale adoption of AI/ML is not an optional curiosity. It is the only way that they can survive, given their scale and business models. There is no way that hundreds of millions of products could be dynamically priced or recommended without using AI/ML.

On the other side are "traditional" businesses, for whom AI/ML has largely been a means to incrementally improve existing processes. For example, banks h ave had statistical and rule-based models in place for credit decisioning and fighting financial crime for decades. They have well-defined frameworks to govern these high-stakes use cases. This can result in a high bar for the adoption of new AI/ML approaches (i.e., "Why break things when they work?"), and limit AI adoption to small pockets of experimentation.

In banking, AI adoption remains low for customer-facing applications.

AI applications for back office functions currently enjoy more traction than those for the front office. Indeed, AI models for claims or underwriting face fewer obstacles than customer-related applications. This may be caused by the perception that customer-facing applications are riskier. For this reason, AI is rarely used in pricing; getting this wrong is considered a major risk.

In contrast, a promising area for AI is conversational intelligence, in which a better understanding of customer needs is achieved through analyzing interactions with banks and insurers. This improved customer understanding can lead to higher customer lifetime value. In line with these risk concerns, it is considered better to start with low risk applications and keep humans in the loop. Once measurable success has been achieved, business leaders can move to medium and finally high-risk applications.

AI is becoming an imperative for companies across industries, even though there is a clear adoption gap between technology-first and "traditional" industries. Some stakeholders have voiced concern about the potential adverse impact of the **European Union Artificial Intelligence ACT** (EU AI ACT) and other emerging AI regulations on AI adoption in Europe and beyond. Attendees of the Executive Roundtables offered a different take.

2. Attitude towards emerging regulations and certification schemes

There is little evidence, yet, of European companies facing a major regulatory handicap in AI adoption.

The existence of a robust privacy regime (GDPR), or the tagging of several AI use cases as “high risk” in the upcoming EU AI Act, do not appear to have yet had a huge impact on European companies’ adoption of AI. There is a fault-line between industries, but our conversations have not unearthed much evidence of a systemic disadvantage arising from being in Europe vs. North America (or Asia). If anything, the certainty that Europe’s regulatory regime and associated standards and certification efforts might bring (vs. the more fragmented and uncertain landscape in the US) could place European companies in a position of advantage.

Some of the companies in geographies without national or regional regulation, such as the United States, welcomed the prospect of regulatory guardrails that would positively channel development efforts and remove bad actors from the field of competition.

> **If anything, the certainty that Europe’s regulatory regime and associated standards and certification efforts might bring… could place European companies in a position of advantage.**

Trustworthy AI cannot (just) be about regulatory compliance.

To date, a lot of the conversation around making AI trustworthy has focused on regulatory expectations around ethics, fairness, transparency, and explainability. However, expecting data scientists and their business or technology partners to meet ethical goals as a standalone objective is futile. The only sustainable way to make AI trustworthy is to ensure such objectives are tied to business KPIs. For example, explainability is arguably even more important for internal buy-in and customer trust than it is for regulatory compliance. Customer and media backlash to perceived or demonstrated malfeasance may be faster, harsher, and more likely than possible regulatory penalties.

The AI Quality Ecosystem is not fully fleshed out.

Indeed, the network of actors in charge of "guaranteeing" AI quality through various tools (e.g., AI quality software) and processes (e.g., testing, audit, and certification) is not fully structured yet. There is no doubt that regulators and standard-setting organisations will play a major role by establishing AI Quality requirements and providing guidance to companies. However, most companies are still unsure about the roles, processes, and tools that they need to implement internally to design, test, deploy and maintain high-quality AI models. There is also uncertainty around the terms governing the relationships between these actors: AI quality software providers, audit/certification bodies, companies and regulators, consumer associations and civil society organisations.

Companies could benefit significantly from AI Quality testing and certification.

Recently, the Hessian Minister of Digital Strategy and Development and VDE (Verband der Elektrotechnik, Elektronik und Informationstechnik, the German Association for Electrical, Electronic, and Information Technologies) announced the establishment of Germany's first AI Quality and Testing Hub. Its aim is to promote AI quality by providing standardization, certification schemes, and testing capabilities to companies using AI systems. Similar initiatives are underway around the world, including in the United States, France, China, and Australia.

Participating in such programs can benefit companies in various ways. First, they will get access to benchmark datasets that could help them compare themselves to their competitors. Second, they will get the opportunity to join working groups where test results are discussed, gaining invaluable information on how to improve their AI models. This is likely to translate into increased revenue and improved customer satisfaction for certified companies.

Focus point

New York City Law on automated employment decision tools

- **There is still a lot of uncertainty around the NYC Law on automated employment decision tools** that requires human resources tech vendors to conduct annual bias audits. One challenge with the law is that it provides very little information regarding how to measure bias and what details should be included in the bias audit report. On reporting specifically, a key challenge is to package information in ways that will not be misinterpreted while remaining understandable for a layperson. Also, its requirement to provide each candidate (internal or external) with 10 business days' notice prior to being subject to the tool seems difficult to implement for high turnover positions such as the ones seen in the retail sector.
- **Yet, the NYC Law and incoming regulations represent a unique opportunity to do it right** because it creates a common baseline that drives good practices. This will push bad actors out of this space and increase customers' trust in compliant vendors. Yet, HR tech providers must act now and implement processes to demonstrate that their models are trustworthy. Part of the answer is increased transparency and education of clients and relevant stakeholders.

In EMEA, business actors are rather welcoming of incoming regulations and conversations have shifted from high-level AI governance frameworks to practical considerations around implementing AI quality and trustworthiness requirements.

3. Elements of a robust AI Quality framework

A broad consensus is forming around core technology requirements for AI quality and Trustworthiness.

Different industries and companies are at varying maturity levels regarding the technical components of a trustworthy AI system. However, a baseline of common requirements is forming, based on guidance from regulators and industry bodies, corresponding standards initiatives, and existing risk frameworks such as model governance in financial services or equipment safety in manufacturing.

A key theme is the emergence of a more holistic "AI Quality" framework, expanding from the previous, individual considerations around explainability, fairness, and ethics.

This new AI Quality framework includes considerations such as:

- Complexity-performance tradeoffs
- Model robustness and resilience in the face of real-world changes
- The security implications of using AI models.

The AI/ML technology ecosystem is fragmented and does not yet address all of these requirements neatly. However, we are seeing rapid evolution in this space, and leading adopters are beginning to orchestrate end-to-end AI/ML pipelines that consciously incorporate AI Quality elements.

Focus point

Not all explainability is alike

- **What constitutes a meaningful explanation is highly contextual.**

  The importance and type of explanation depends on the use-case being considered. For instance, in operations planning or predictive maintenance, explainable AI is essential to investigate when things go wrong and fix the system in a timely fashion. When it comes to facial recognition, a major risk is having human agents uncritically following the system's recommendations. Here, explanations are rather needed to demonstrate that decisions (e.g., arresting a suspect) were not fully automated. Yet, there is a consensus over the fact that the higher the criticality of the use-case, the more important is the explanation in order to increase trust in AI models.

Fairness and performance are both part of AI Quality.

Sometimes, there is a perception that increasing fairness would lead to a decrease in performance. Yet, it does not need to be a zero-sum game. Indeed, you can actually optimize both with appropriate tools and leadership.

Lack of a common standard for AI Quality reports.

Data science teams must provide detailed documentation of the functioning of AI models across their lifecycle. This will:

- Drive the adoption of AI-powered applications
- Achieve internal buy-in from management and business leaders
- Demonstrate compliance with emerging regulations..

However, there is no standard documentation framework for AI Quality reports. In banking, this situation leads to a constant back and forth between model risk and data science teams to refine existing model risk management frameworks. This process is very manual and ineffective. One key dimension of this challenge is that model risk management teams are asking for qualitative documentation, explaining the logic of all AI models.

4. Steps to scale Trustworthy AI

A thoughtful AI Quality framework will have little impact if one fails to build the appropriate organisational capabilities to operationalize it. To address this challenge, leading AI adopters offer the following process as a starting point:

Make sure that you all speak the same language internally.

One of the key challenges when it comes to operationalizing trustworthy AI is to close the communication gap between domain experts and technical teams. Indeed, some of the concepts related to trustworthy AI (e.g., fairness metrics) may mean different things for data science teams and business leaders. Also, there are different levels of understanding of a model's behavior, depending on the perspective being adopted. Therefore, it is important to define trustworthy AI requirements at the company level and share information across departments on the characteristics of AI models currently being used.

Define clear lines of accountability.

At the moment, most organisations don't have model validation processes to assess their ML models against established fairness metrics. This is the case even in industries that have well-established frameworks around Model Risk and Data Management, such as banking and healthcare. There are organisations where executives feel confident about their ability to mitigate bias with various tools, yet they struggle to act decisively because no one owns this risk. As a result, bias mitigation ends up being an afterthought. Part of this accountability falls on the developers of AI models but this should be clearly reflected in their mandate and they can't act alone. Therefore, someone should be in charge of this at the management level.

Take leadership and empower employees.

Technical skills only account for 30% of the overall successful deployment of AI models. Strong leadership is needed to get the buy-in of business leaders, to select those few projects where AI can add value in a short term, and to establish clear milestones during project development and execution. In addition, training programs are essential to driving the adoption of AI-based applications. Also, data science teams need appropriate tools to build machine learning pipelines and ensure model transparency despite increasing model complexity. Finally, strengthen organisational capabilities through training and closer cross-functional collaboration, particularly between data science and model governance teams.

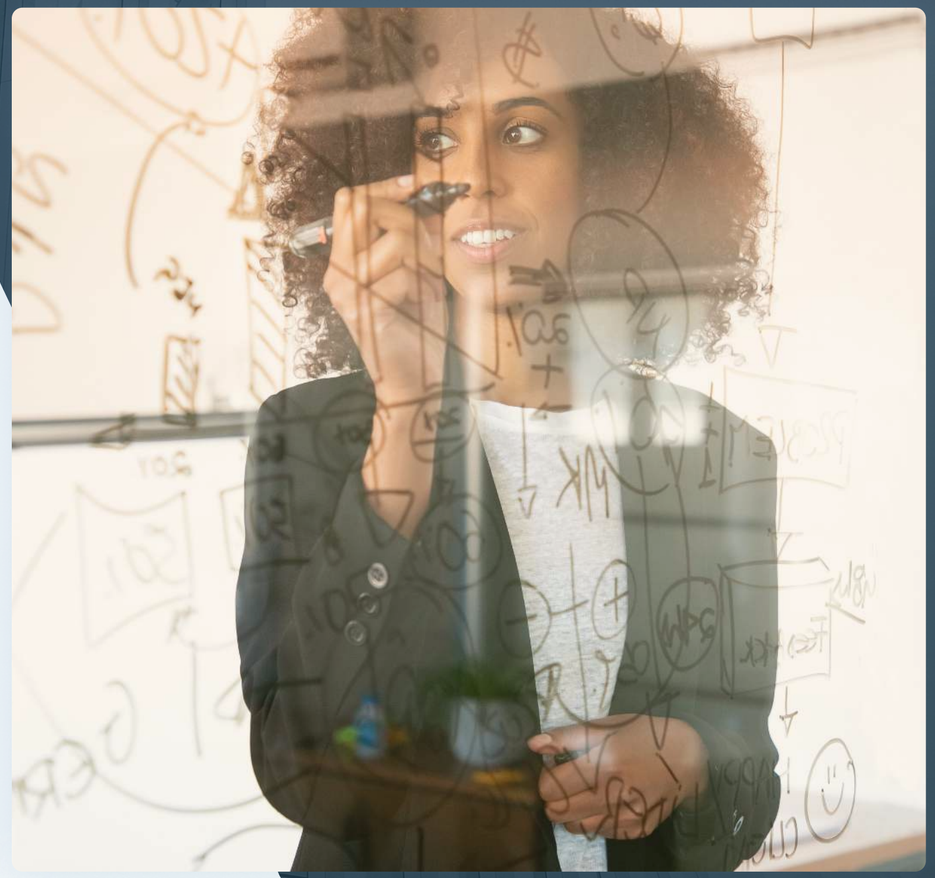
Continuously test and monitor AI models.

AI models evolve with data and use. As a result, their behaviors are hard to anticipate. Thus, it is essential to continuously test and monitor AI models to guarantee that their behaviors are consistent with a set of fundamental truths. This process will help identify the main determinants of specific AI models, as well as identify those edge cases where they are likely to fail. Adding feedback mechanisms for user communities and A/B testing of various versions of the same model are also meaningful ways to improve trustworthiness.

Conclusion

These discussions provided valuable insights into industry actors' responses to the trustworthy AI challenge. Their focus has clearly moved from high-level discussions around Trustworthy AI to practical considerations around implementing AI quality requirements into their model development, testing, and deployment processes. Most are still unsure about how to do this at scale, but leading adopters are showing the way. They demonstrate strong leadership from top management, facilitate cross-functional collaboration, establish clear lines of accountability, and provide appropriate tools to data scientists.

Further, across the roundtables, there was a sober optimism about their ability to complete this journey, partly based on the sense that the gap between regulatory and technology requirements is rapidly narrowing. The AI/ML technology ecosystem is still fragmented, but some tools are well identified. Regulatory uncertainty persists in some areas but customer demand for trusted AI systems is driving action.



Lead author

Lofred Madzou
Director of Strategy and Business Development
TruEra