

EXPLAINED
EXPLAINABLE AI



The huge opportunities and benefits that AI offers don't come risk free, of course. What kind of innovation does? But it certainly pays to begin an AI journey with a clear-sighted view of what the risks might be for an organisation.

So, what are they? We think there are four principal risks that must be considered up front. These relate to trust, liability, security, and control:

TRUST

How do we demonstrate to citizens that an AI is safe to use? How do we avoid biases, unconscious or not, being written in from the outset? The answers to these questions lie in transparency and accountability. Decisions taken by an AI must be open to appeal and interrogation.

LIABILITY


What happens when an AI makes an error – or even breaks the law? Who is legally responsible? Changes to legislative and regulatory requirements will need to be monitored carefully.

SECURITY

How do we prevent unauthorised or malicious manipulation of an AI? Security becomes paramount, and is compounded by the increasing use of open source code.

CONTROL

What happens when a machine takes over a process? How does a human take it back if they need to? Careful thought is needed about when and how control is transferred between humans and AIs. For example, it is all very well providing a human riding in a self-driving car with the means to take control but if they are not paying attention 100% of the time, they will be unlikely to intervene fast enough in a critical situation.

A large iceberg floats in a blue ocean. The tip of the iceberg is visible above the water surface, while the much larger, jagged base is submerged underwater. The water is a deep blue, and the sky is a lighter blue with some clouds. The overall scene is a metaphor for hidden or unseen aspects of a business or organization.

**72% of executives
report that their
organisations seek
to gain customer
trust and
confidence by
being transparent
in their AI-based
desicions and
actions.¹**

Accenture Technology Vision,
2018

Accenture Launched “Pinterest For AI Education”, a new technology platform to train more than 180,000 of its employees globally in the latest digital technologies in just over 20 months. It now plans to use the interactive platform with clients to help develop their IT workforces in critical areas such as digital, cloud, security and artificial intelligence. The Accenture Future Talent Platform integrates learning services and curriculum on as-a-service and mobile platforms to help workforces move away from traditional training and foster a culture of continuous learning.²

These issues are something we’ve given a great deal of thought to at Accenture. Our strong recommendation is to take a “human first” approach to AI thinking. And that means adopting a framework for what we call “Responsible AI”. This framework recommends mitigating the risks of using AI with four imperatives: govern; design; monitor; and reskill:

01. GOVERN

Create the right governance framework for AI to flourish. Anchor it to your organisation's core values, ethical guardrails, and accountability frameworks.

02. DESIGN

Build trust into your AI from the outset by accounting for privacy, transparency, and security from the earliest design stage.

03. MONITOR

Audit the performance of your AI against a set of key metrics. Make sure algorithmic accountability, bias, and security metrics are included.

04. RESKILL

Democratise the understanding of AI across your organisation to break down barriers for individuals impacted by the technology.

It's also vital to remember that humans can be susceptible to unconscious bias. And that has big implications when it comes to coding and training an AI and selecting the data sets it will use. This is an area in which general standards may be required. Collectively, we may need to aspire to a higher degree of responsibility from our AIs than we would necessarily demand from human colleagues.

The need to explain

There is one thing above all that will ensure public trust is maintained when an organisation starts using AI: “explainability”. In other words, being ready to explain how and why an AI came to the decision it did. This is something that certain regulated industries are already familiar with. Financial services institutions, for example, are required to explain the decisions they take that affect their customers.

But there’s also a broader issue here: humans are more likely to trust something they understand. So “explainable AI” becomes a vital part of any AI strategy.

DARPA’S EXPLAINABLE AI CHALLENGE

DARPA, the Defense Advanced Research Projects Agency, has launched an explainable AI (XAI) programme with the goal of creating a suite of machine learning techniques which produce more explainable models. The models will be combined with interfaces capable of translating them into understandable and useful explanations for human users.³

In fact, that's easier said than done. Machine learning is often by its very nature a "black box" exercise. In other words, it operates in ways that can make it very hard to explain how it arrived at the outputs it produced. But many AI practitioners and data scientists are thinking about this question, and new approaches offering better explanations of the science underlying AI decisions will likely soon emerge.

In the meantime, there are some practical steps that every business can take now to make their AI more explainable:

01. INVENTORY

Think about the decisions that are or will be taken by AI in your organisation. Which of them would require an explanation – or create an expectation of one? Do they relate, even indirectly, to key areas like employment, recruitment, lending, education, healthcare, housing, inclusion, or safety?

02. ASSESS

Consider any quantitative and qualitative models that are already providing explanations for decisions taken by AI. How are they performing for their intended recipients?

03. DESIGN

Revisit the design principles used for your AI. How could they make the process of making decisions more human-centred and understandable?

04. AUDIT

Review the data. How do you ensure your AI is using data sets that reflect the evolving nature of your workplace?

Recommended Reading

The Second Machine Age by Erik Brynjolfsson and Andrew McAfee

Machine, Platform, Crowd by Erik Brynjolfsson and Andrew McAfee

Life 3.0 by Max Tegmark

Human + Machine: Reimagining Work in the Age of AI by Paul Daugherty & James Wilson
(forthcoming March 20th, 2018)

Homo Deus by Yuval Noah Harari

The Quest for Artificial Intelligence by Nils Nilsson

The Master Algorithm by Pedro Domingos

The Future of the Mind by Michio Kaku

I would like to thank Accenture colleagues too numerous to mention for their help formulating the concepts described in this publication; and Lucy Frost, Noor Sajid, Caryn Tan and Alexandra Vernon for their assistance with research and design.

Sources

- ¹ Accenture Technology Vision. (2018). Redefine Your Company Based on the Company You Keep: Intelligent Enterprise Unleashed. https://www.accenture.com/t20180208T172438Z_w_/us-en/_acmedia/Accenture/next-gen-7/tech-vision-2018/pdf/Accenture-TechVision-2018-Tech-Trends-Report.pdf
- ² Which.(2017, November 20) Accenture Launches Pinterest for Education. Which-50. <https://which-50.com/accenture-launches-pinterest-ai-education/>
- ³ "Gunning, G.(n.d) Explainable Artificial Intelligence. DARPA. <https://www.darpa.mil/program/explainable-artificial-intelligence>"usts-its-prices-2015-1

About Accenture

Accenture is a leading global professional services company, providing a broad range of services and solutions in strategy, consulting, digital, technology and operations. Combining unmatched experience and specialized skills across more than 40 industries and all business functions – underpinned by the world’s largest delivery network – Accenture works at the intersection of business and technology to help clients improve their performance and create sustainable value for their stakeholders. With approximately 449,000 people serving clients in more than 120 countries, Accenture drives innovation to improve the way the world works and lives.

Visit us at www.accenture.com.

Applied Intelligence

Accenture Applied Intelligence, part of Accenture Digital, applies artificial intelligence (AI) and human ingenuity at the core of business to help clients become intelligent enterprises and solve their most complex business problems. By deploying AI responsibly and combining it with our deep industry and analytics expertise, we enable the digital transformation of organisations, extend human capabilities, and make intelligent products and services a reality.

Follow [@AccentureAI](https://twitter.com/AccentureAI) and visit accenture.com/appliedintelligence.

Copyright © 2018 Accenture
All rights reserved.

Accenture, its logo, and High Performance
Delivered are trademarks of Accenture.