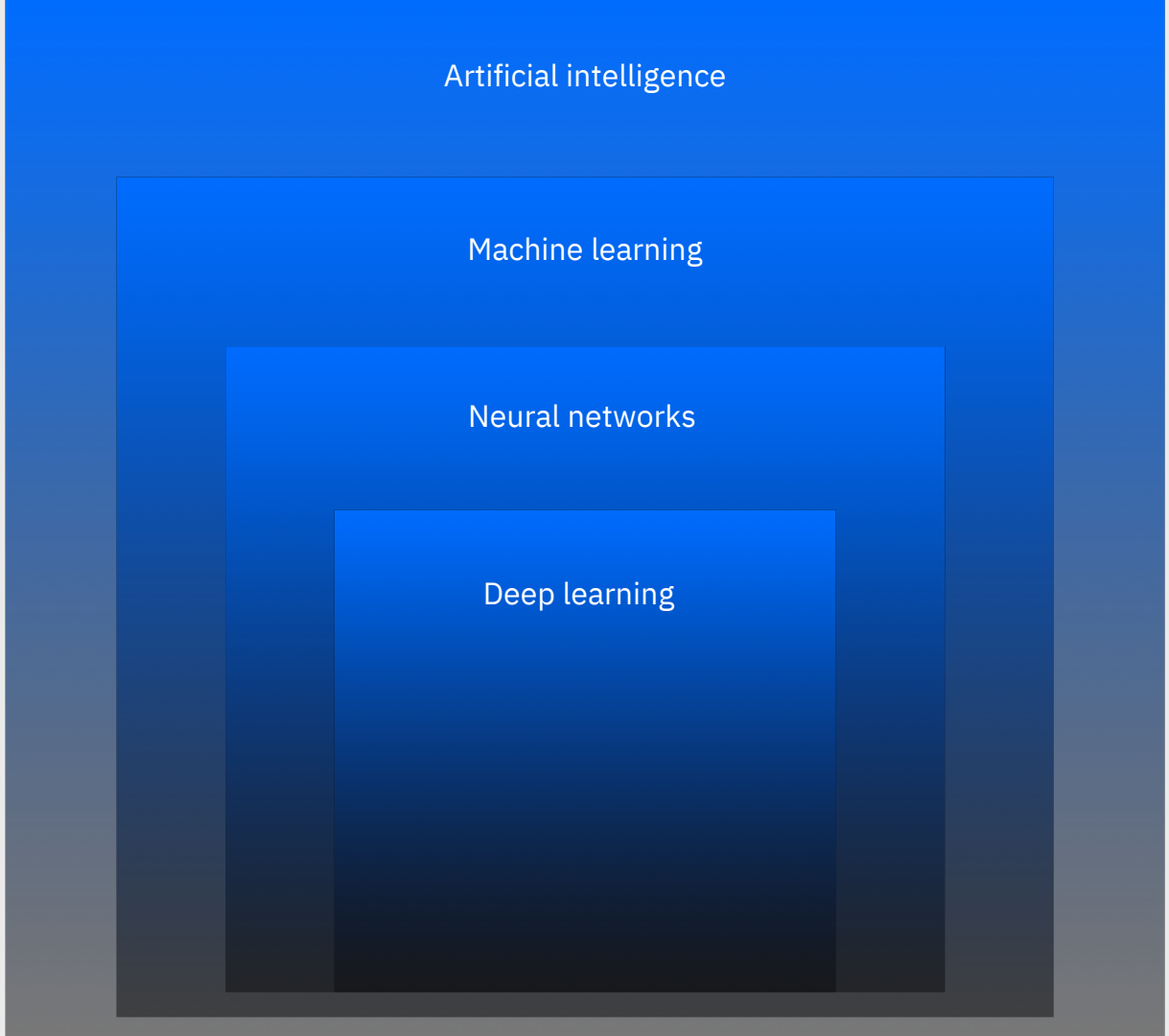
Trusted AI for Business

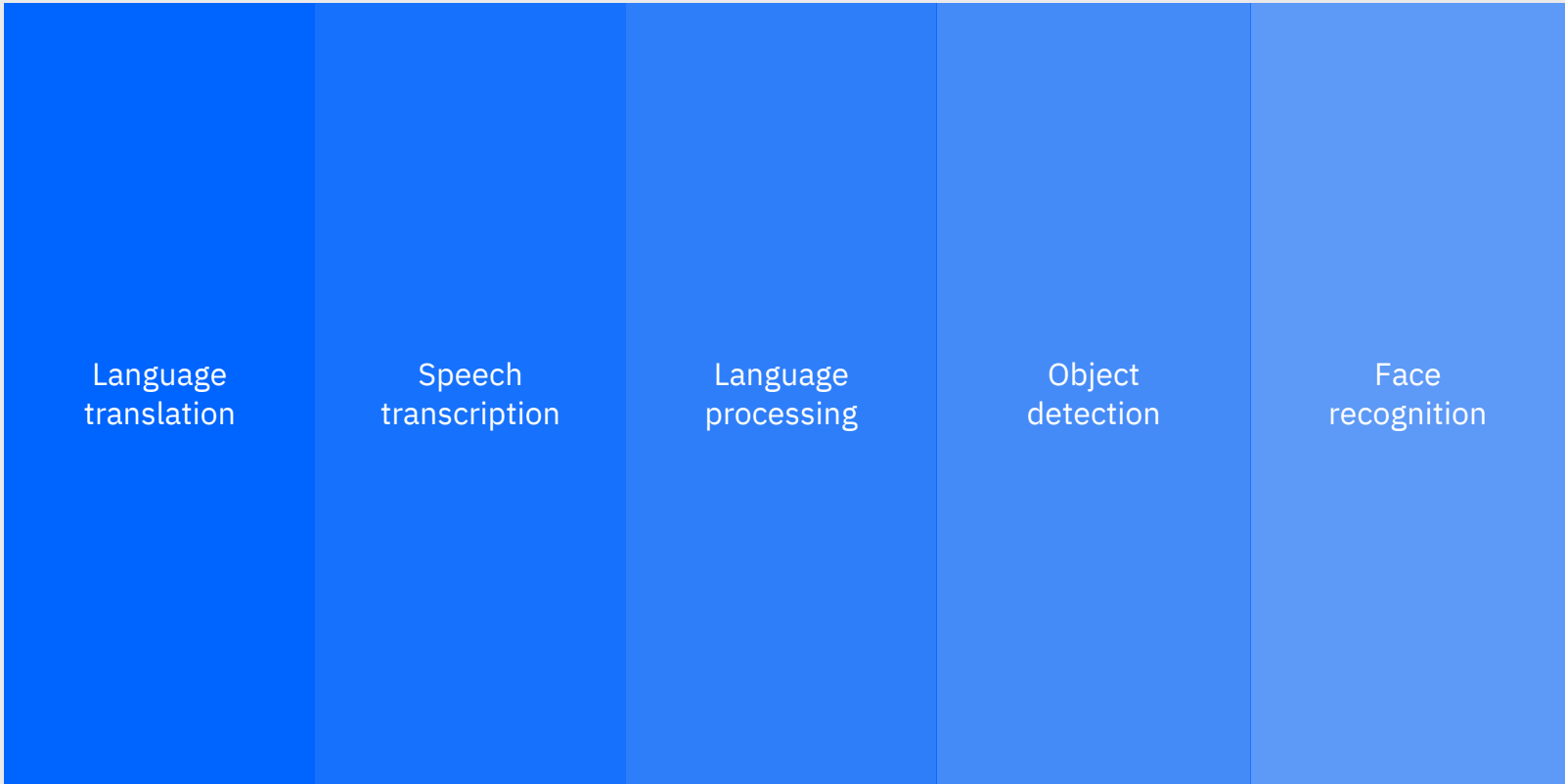
Presented by: **Alessandro Curioni, IBM Fellow, Vice President Europe and Director, IBM Research Zurich**

For: **ETSI Summit on Artificial Intelligence**

AI is...


Artificial intelligence
Machine learning
Neural networks
Deep learning

Specialized AI works incredibly well



94%

of companies believe that AI is the key to competitive advantage

80%

of data is either inaccessible, untrusted or unanalyzed

81%

do not yet understand the data required for AI

1 in 20

companies have extensively incorporated AI into offerings and processes

60%

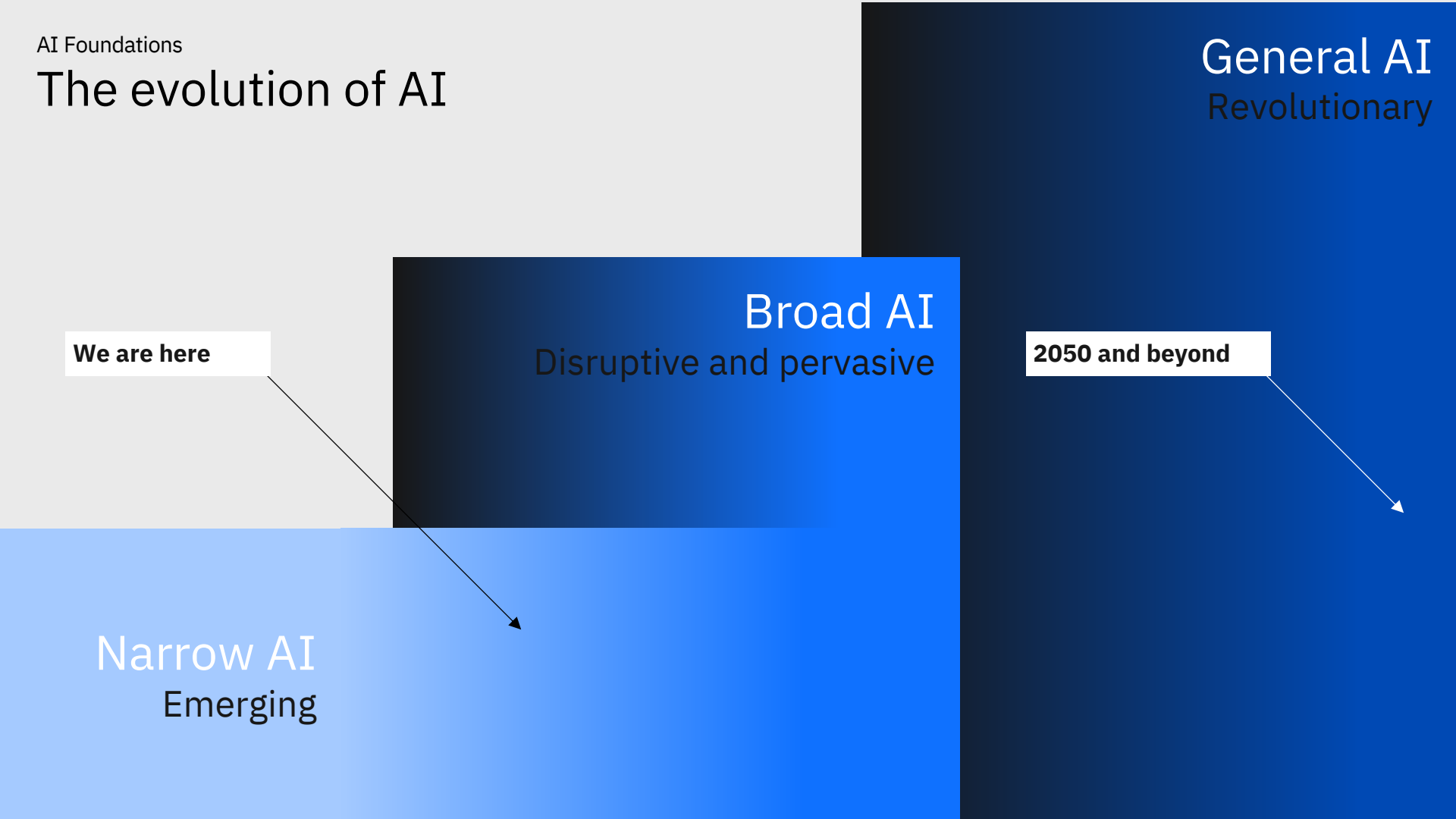
See compliance as a barrier due to a lack of trust in AI outcomes

65%

do not fully trust their own organizations analytics

AI Foundations

The evolution of AI



The evolution of AI

Narrow AI

Deep learning

Single-task, single-domain, with superhuman accuracy

Requires large amounts of labeled data

Broad AI

Learning + reasoning

Multi-task, multi-domain, multi-modal

Learns with much less data

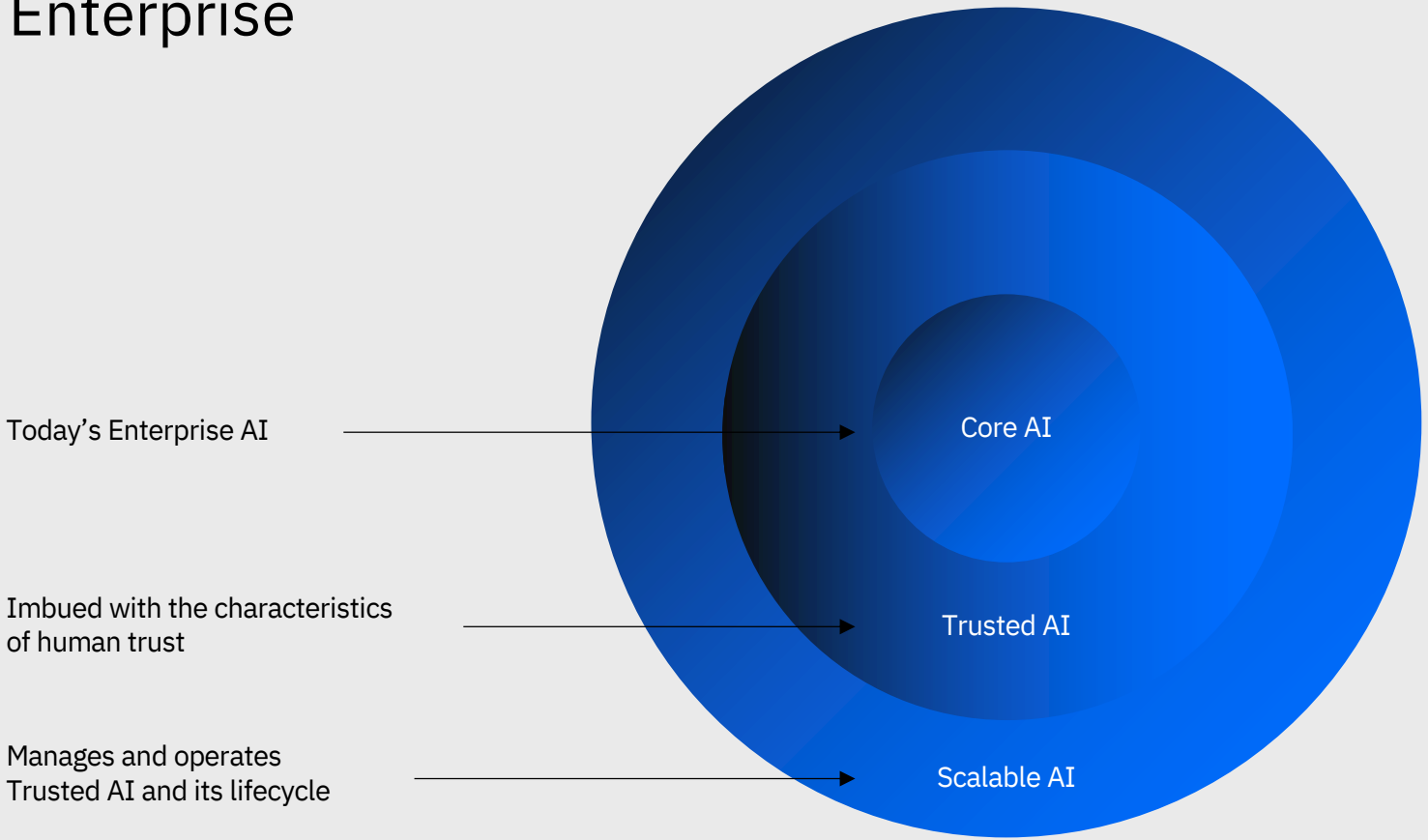
General AI

True neuro-AI

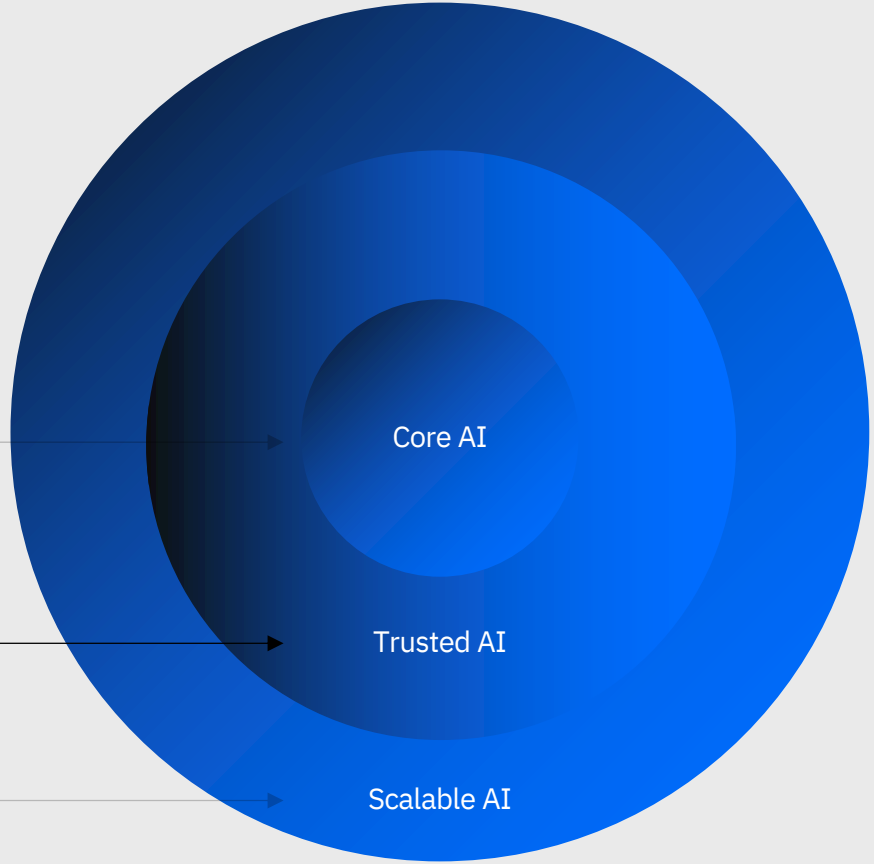
Cross-domain learning and reasoning

Broad autonomy

AI for Enterprise



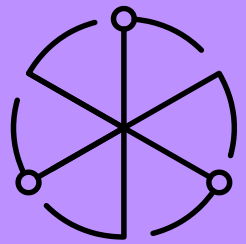
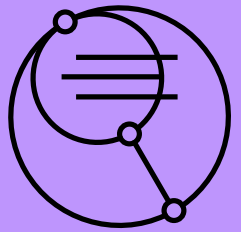
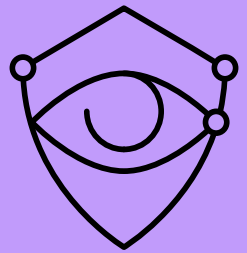
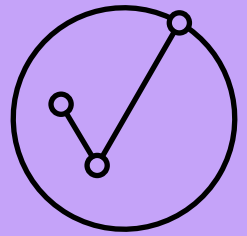
AI for Enterprise



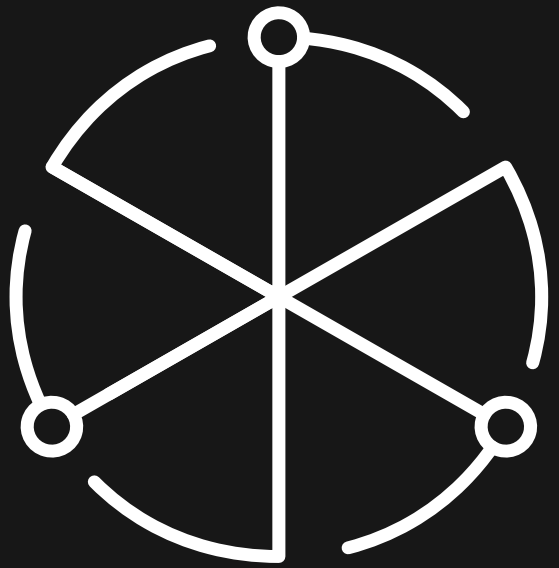
What does it take to trust a decision made by a machine?

Trusted AI

The pillars of trust

Is it fair?	Is it easy to understand?	Is it secure?	Is it accountable?
			
Fairness	Explainability	Adversarial Robustness	Transparency

Making AI Fair

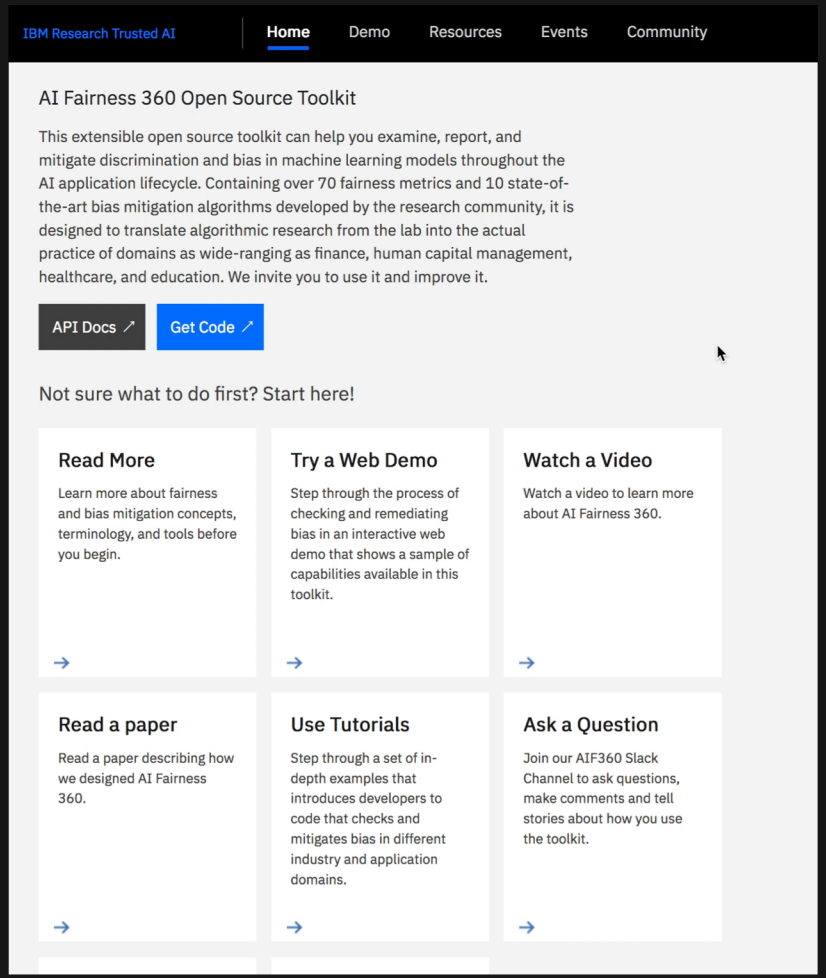


Trusted AI

AI Fairness 360:

- 30+ fairness metrics/checkers
- 10 bias “mitigators”
- industry tutorials

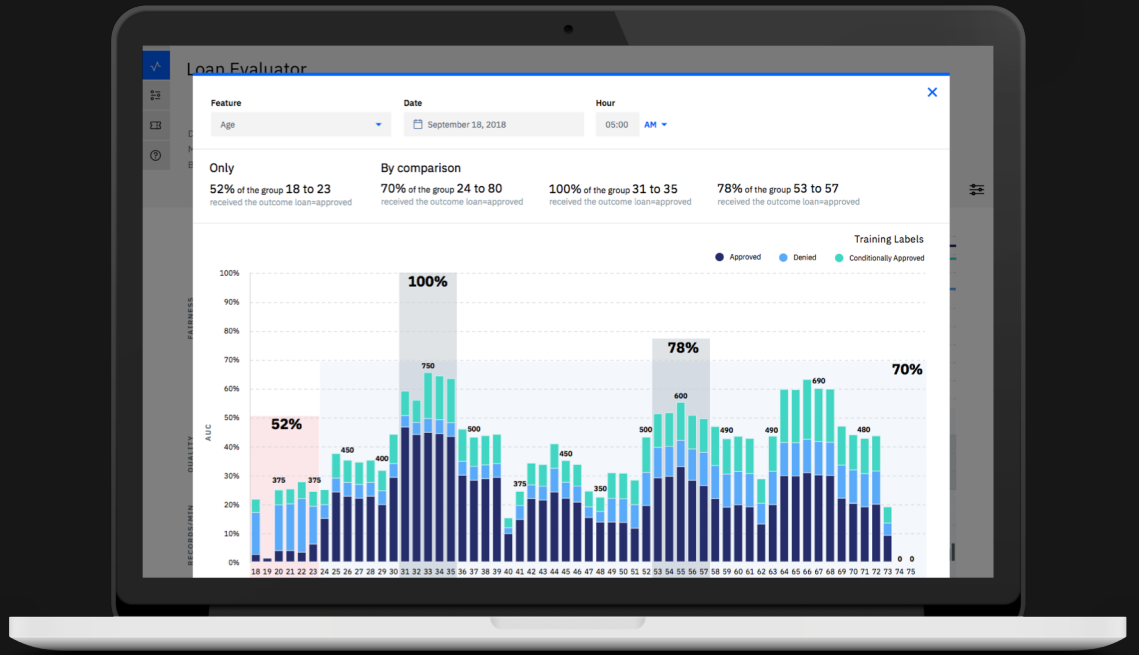
aif360.mybluemix.net/



Trusted AI

Fairness in action: Watson OpenScale

Operationalizes and instruments fairness into enterprise-grade workloads



Making AI Explainable

Algorithm: Contrastive explanations

We deduce that the image is “3” because,

we observe that curves are present,

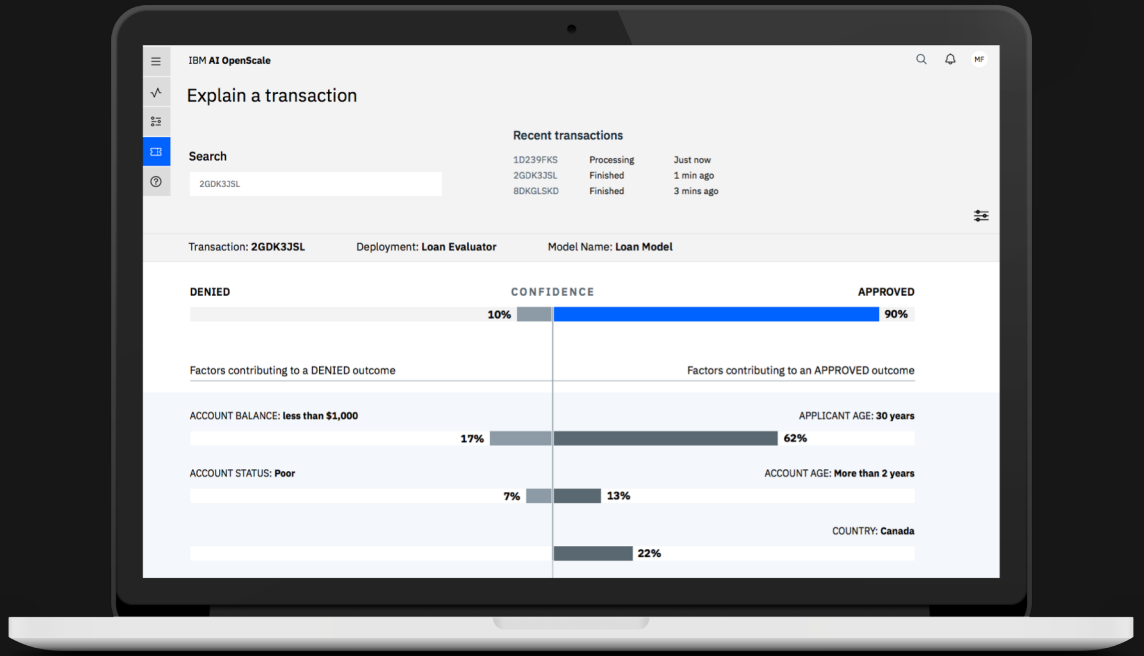
but a dash is absent.

Dhurandhar et al, **“Explanations Based on the Missing: Towards Contrastive Explanations with Pertinent Negatives,”** NeurIPS 2018.

Trusted AI

Explainability in action: Watson OpenScale

Instruments explanations into enterprise-grade workloads

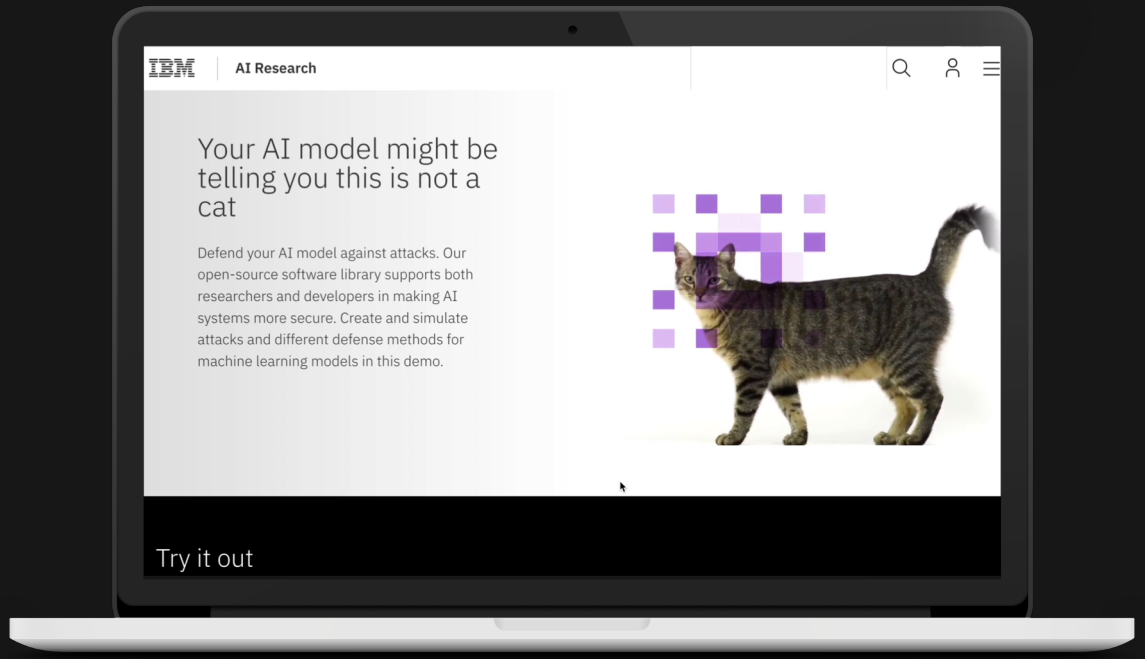


Making AI Secure

Trusted AI

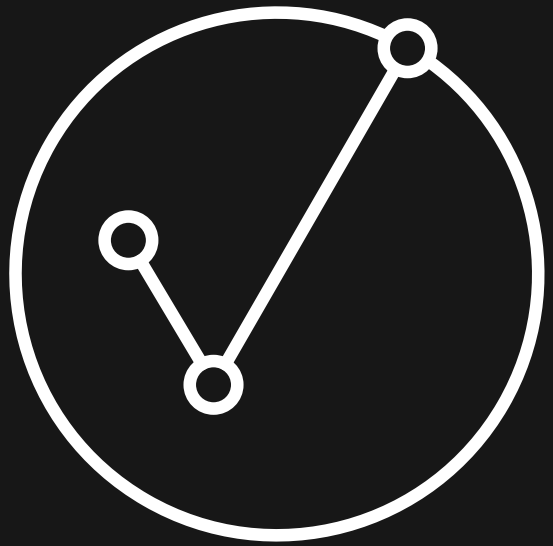
Adversarial Robustness Toolkit

The most comprehensive toolkit for “attacking” and defending AI



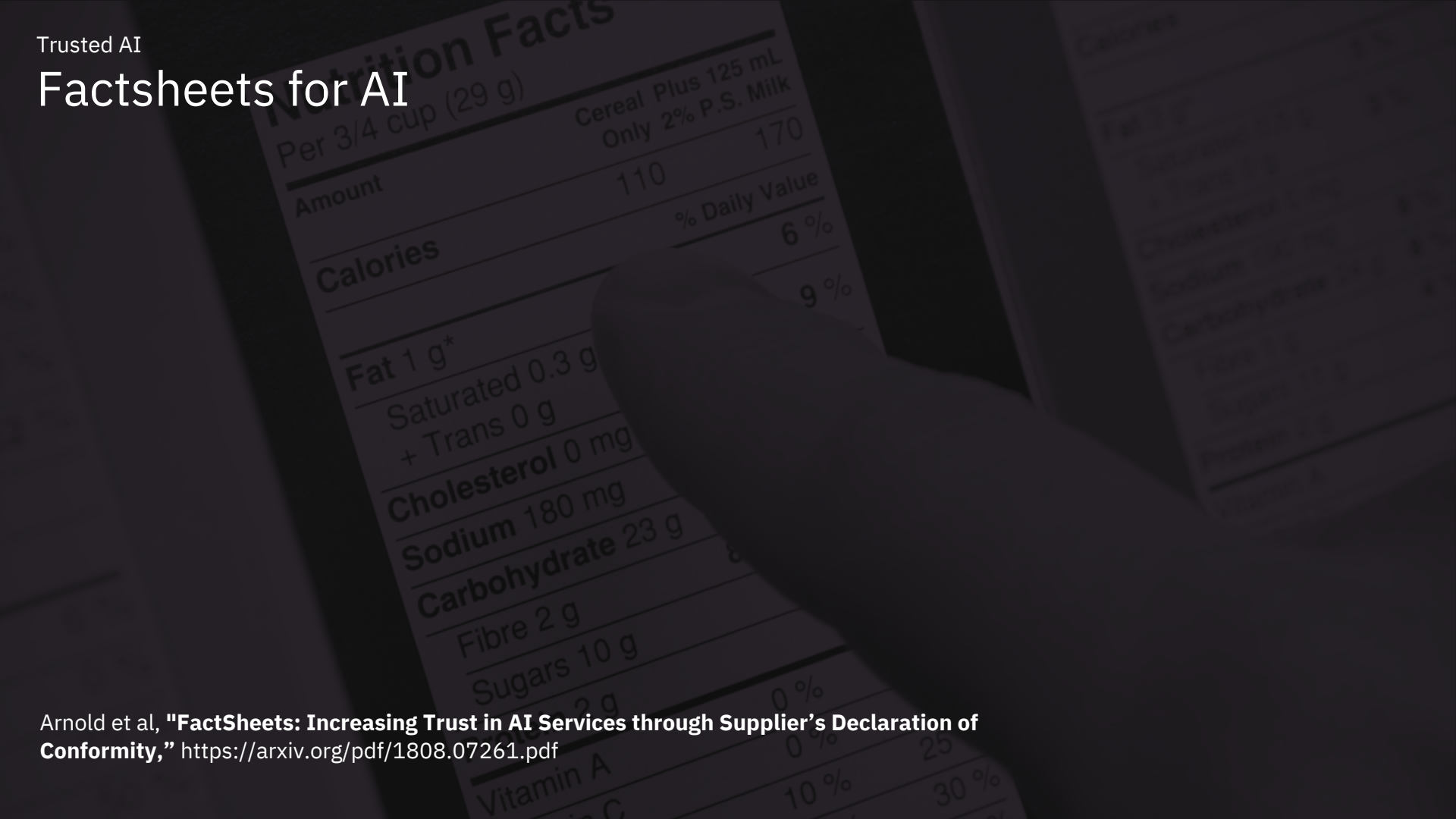
https://art-demo.mybluemix.net/

Making AI Transparent



Trusted AI

Factsheets for AI

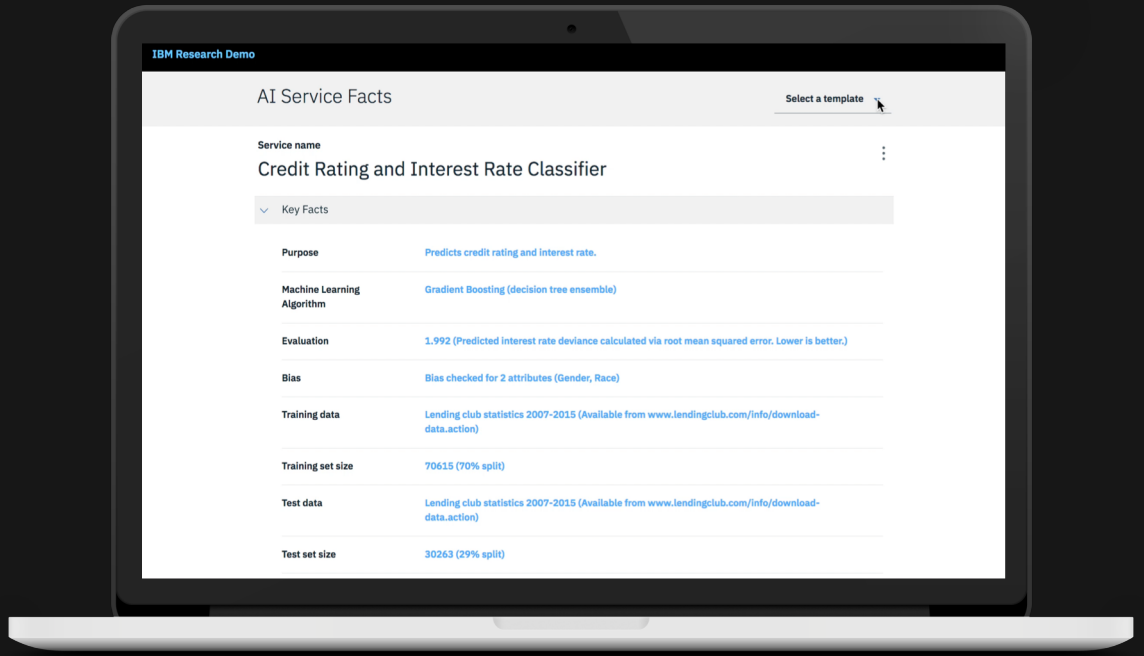


Arnold et al, **"FactSheets: Increasing Trust in AI Services through Supplier's Declaration of Conformity,"** https://arxiv.org/pdf/1808.07261.pdf

Trusted AI

Transparency in action: Factsheets for AI

A transparent reporting mechanism for how an AI system operates and performs



Trusted AI

Coming Soon

Ethics guidelines for trustworthy AI

AI Ethics Guidelines presented by the European Commission's High-Level Expert Group on Artificial Intelligence (AI HLEG).

Building more trustworthy AI systems is not only a research question, it’s a business imperative.