WELCOME TO

AI GOVERNANCE

AI for Governance and Governance of AI

BCS IRMA January 12th, 2021

Mike Small,
Senior Analyst | Kuppingercole Analysts AG

Agenda

01 What is AI

02 AI for governance

03 Governance of AI - Example Scenario

04 AI Governance Frameworks

AI is made up of many disciplines

Artificial Intelligence is an umbrella term that includes multiple different technologies

Natural Language Processing

A computer's ability to extract meaning and information from written and audible speech. Includes natural language generation and natural language understanding.

Machine Learning

An algorithm that alters itself over time as it is exposed to data so that it "learns" and thus improves itself without being explicitly programmed. Includes supervised learning, unsupervised learning, and deep learning/Neural Networks.

Machine Reasoning

An autonomous agent's ability to reason with knowledge so as to plan strategies and carry out action sequences. Includes expert systems and planning/scheduling/optimization.

Robotics

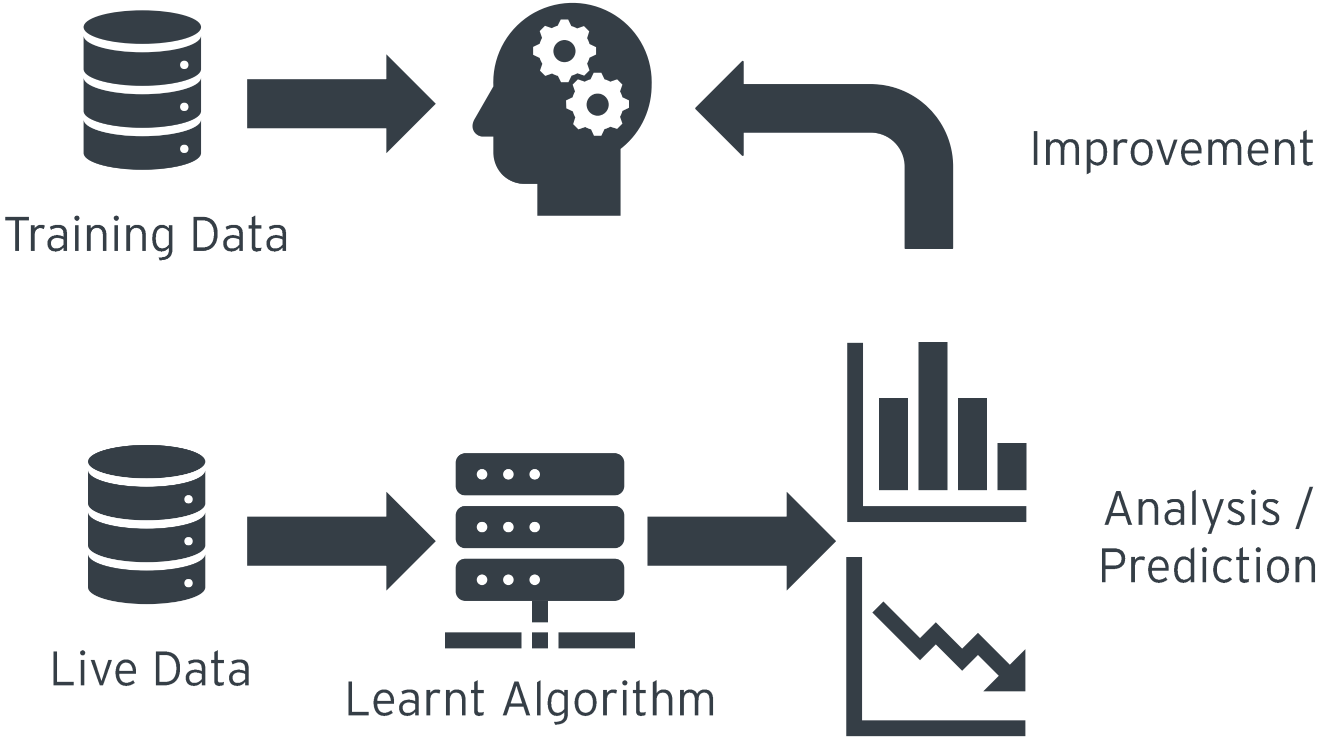
Programmable machines that carry out tasks (semi-)autonomously. AI can enhance robotics to produce autonomous vehicles, autonomous agricultural equipment, surgical assistance, etc.

Computer Vision

A computer's ability to process image, video, and live feeds. Includes optical character recognition, image recognition/classification, and facial recognition.

Machine Learning

A completely new development process



Training Data

Improvement

Live Data

Learnt Algorithm

Analysis / Prediction

Hype vs. Reality in AI & ML: Where are the Concrete Business Benefits?

Agenda

01 What is AI

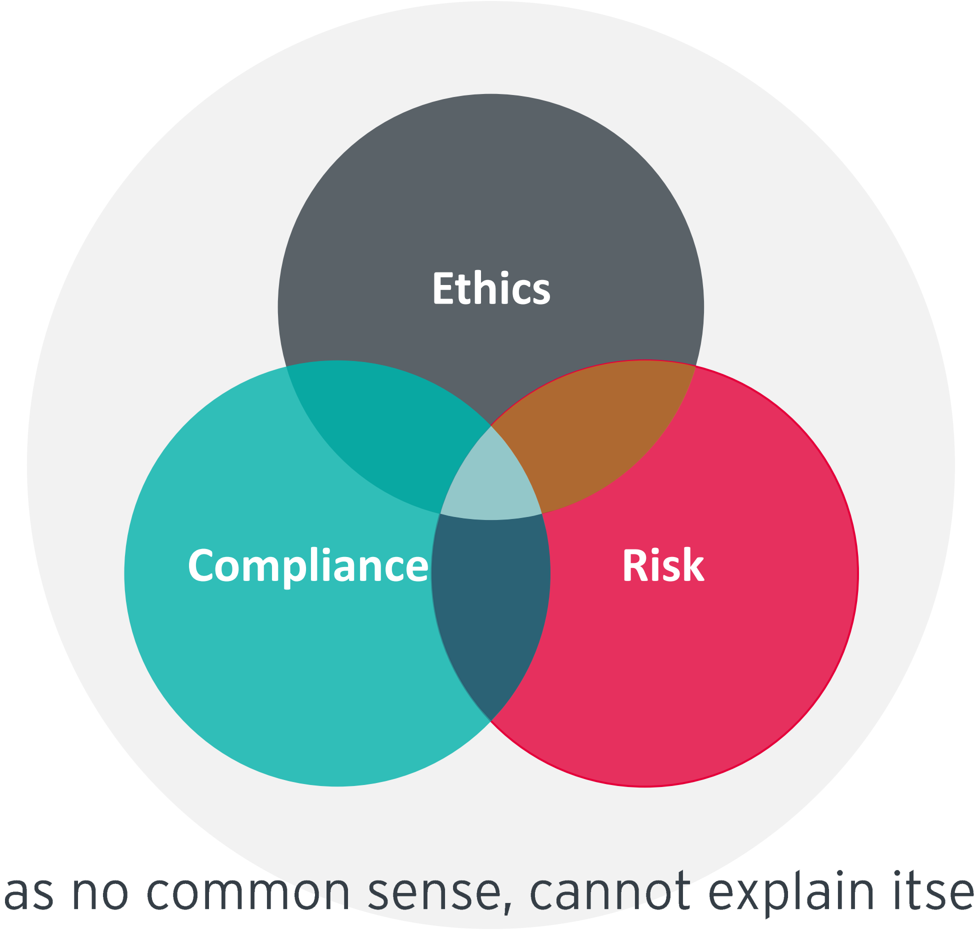
02 AI for governance

03 Governance of AI - Example Scenario

04 AI Governance Frameworks

Governance

Sets objectives and boundaries for execution



AI has no common sense, cannot explain itself, and is not responsible for its actions

Governance Problems

Where AI could help

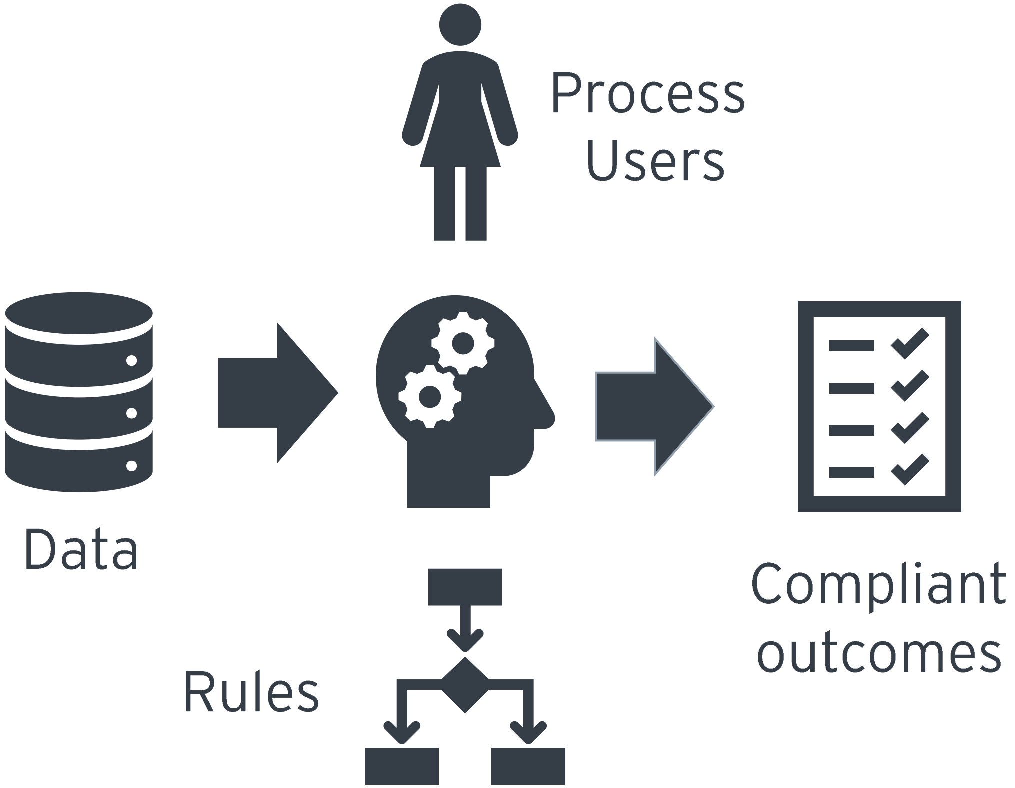
- ⊗ Making sense of the vast amount of data
- ⊗ Tidal Wave of Regulation
- ⊗ Lack of skills
- ⊗ Is AI a friend or a foe?

kuppingercole ANALYSTS

Process Assistant

Support for complex processes with compliance impact

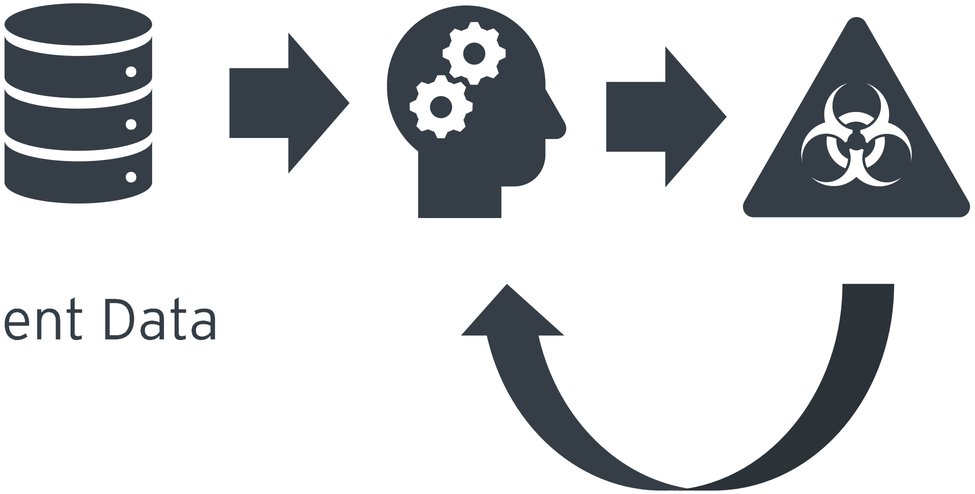
- ☒ Assists with complex rule based processes.
- ☒ Helps the user to navigate complex rules.
- ☒ Reduces risks of inadvertent non-compliance and saves costs.



Categorizing Event Data

User and Entity Behaviour Analytics

- Now widely integrated into security tools.
- Training period for the system to learn what is “Normal”.
- Identifies “abnormal” behaviour and feedback tunes out errors.



Event Data

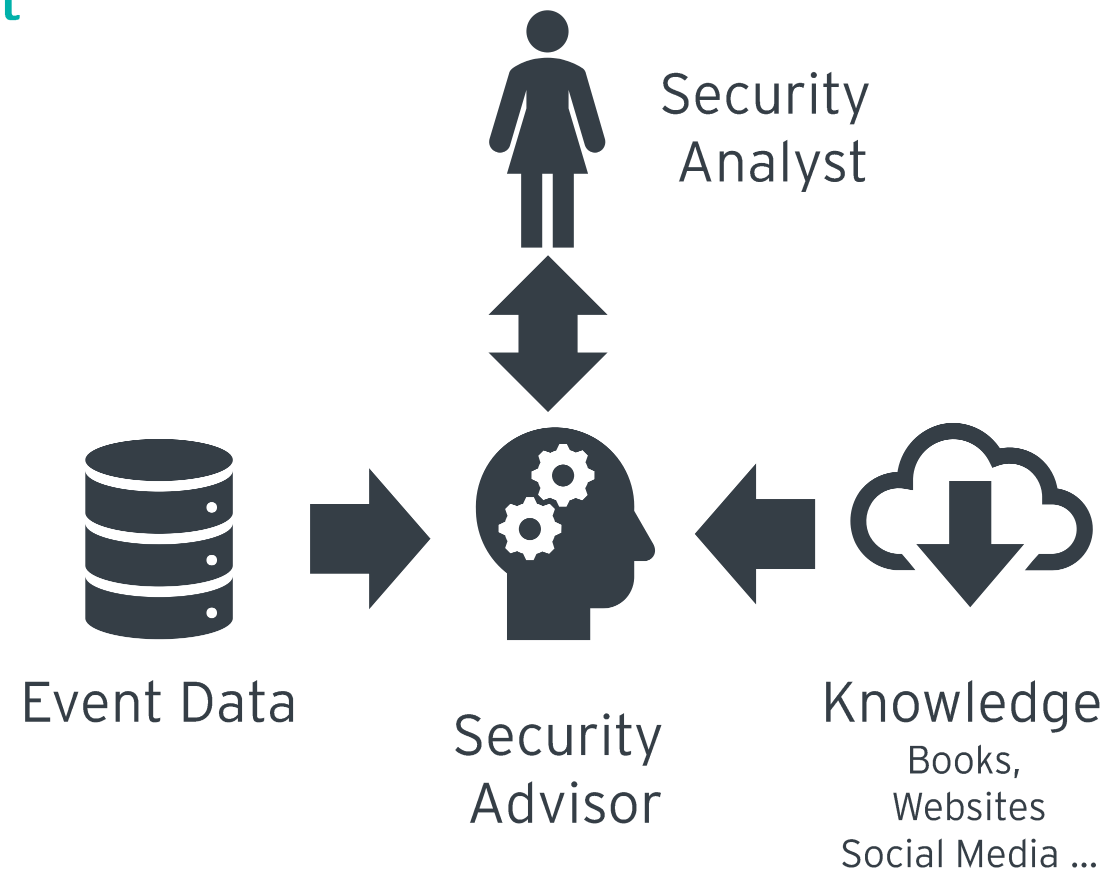
Detects what is Abnormal and learns from mistakes

The Role of AI in Modern Business Intelligence Platforms

The Intelligent Assistant

Assists rather than replaces the security analyst

- ⊗ Ingests natural language threat data from manuals, textbooks and social media sites
- ⊗ Identifies event anomalies
- ⊗ Relates the anomalies to the threat data to accelerate diagnosis and remediation

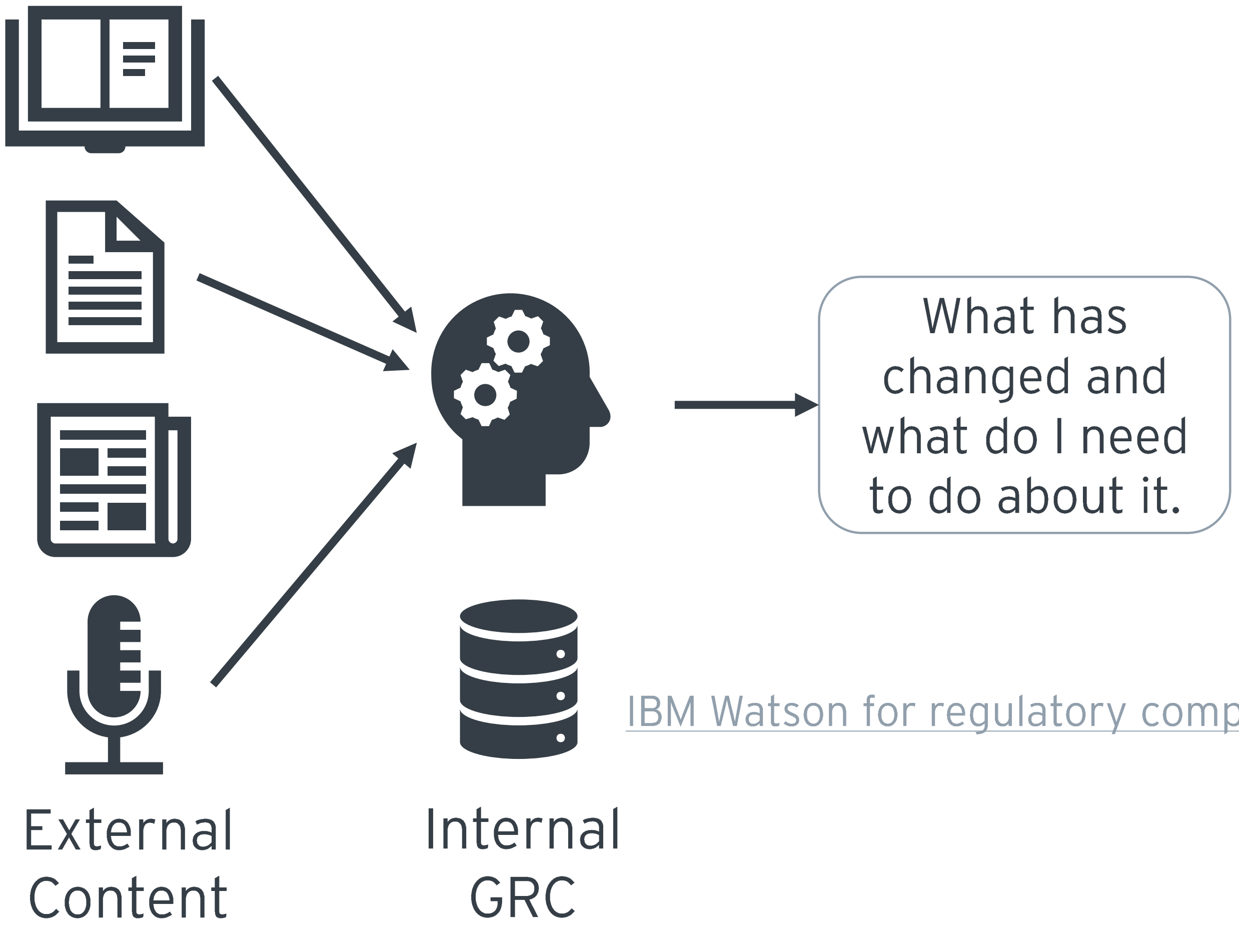


IBM QRadar Security Advisor

The Compliance Assistant

Helps organizations to manage compliance with emerging regulations

- Ingest of regulations to identify the obligations.
- Compare obligations against controls required.
- Correlate this with other regulations and personalized data about the organization.



IBM Watson for regulatory complianc

Agenda

01 What is AI

02 AI for governance

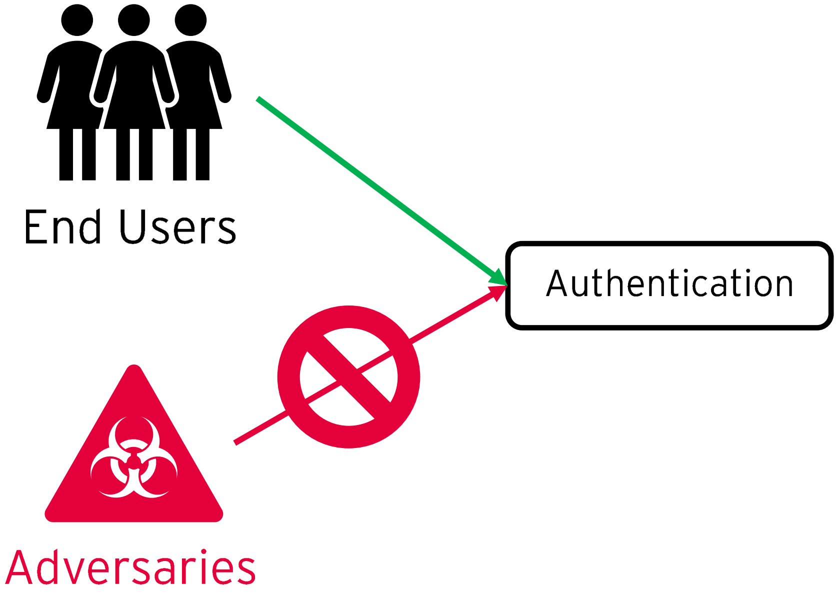
03 Governance of AI - Example Scenario

04 AI Governance Frameworks

Scenario

An example to illustrate the governance challenges

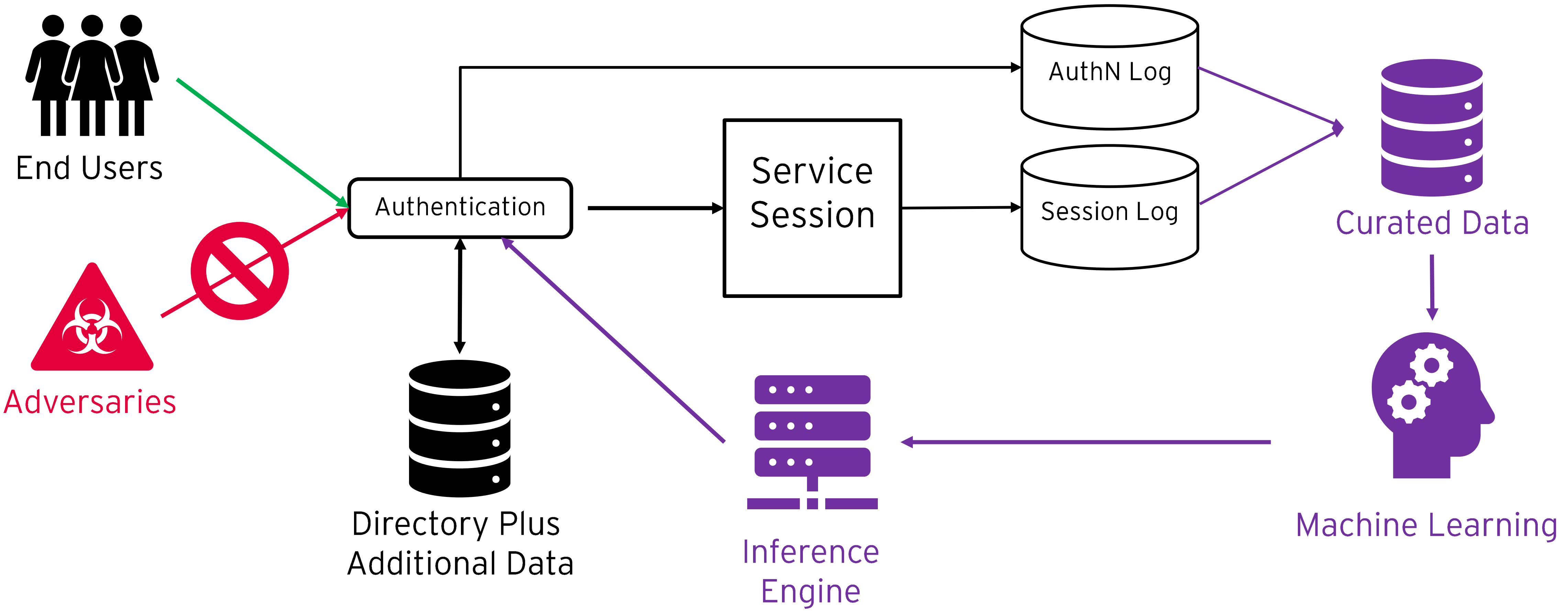
How could machine learning be used to improve user authentication?



Based on workshop from cybernetix.world 2020 – KuppingerCole

The Concept

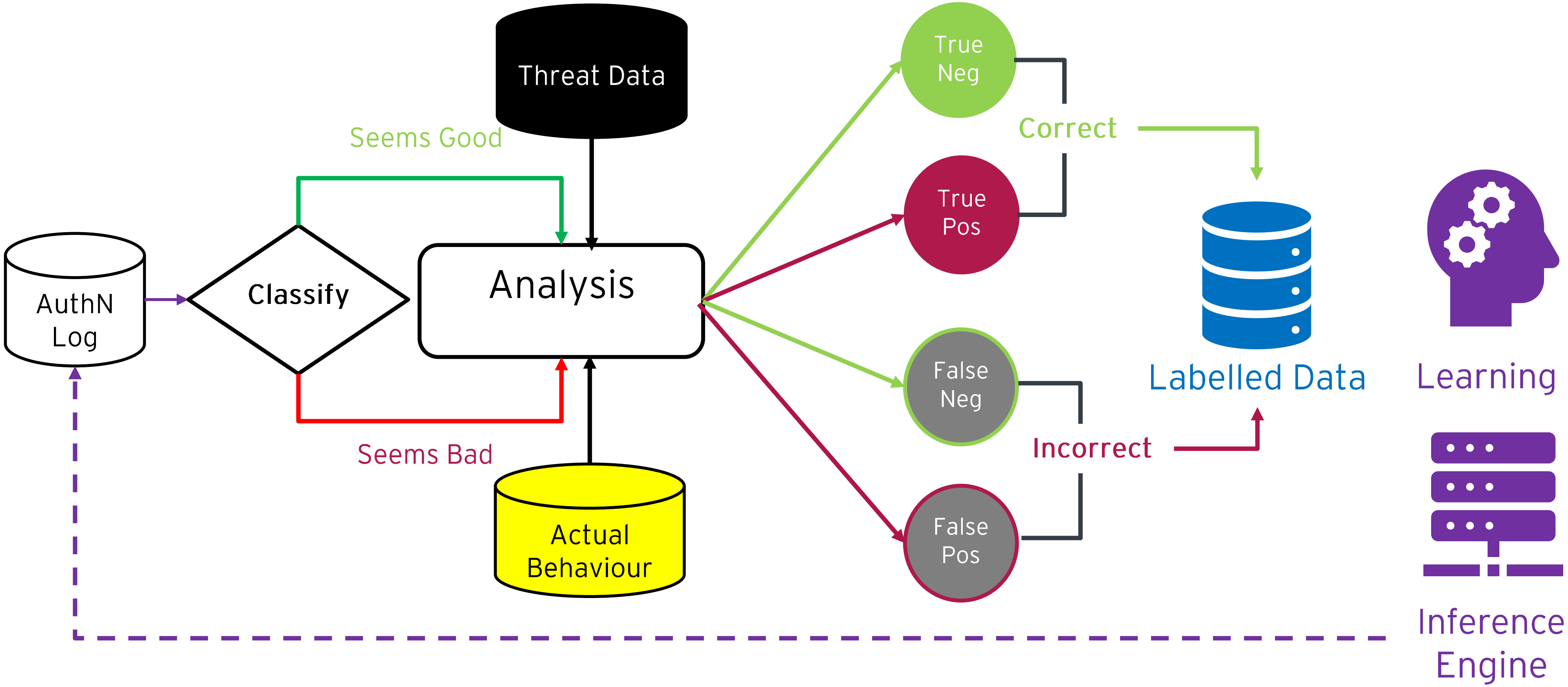
To apply machine learning to improve the accuracy of authentication



© KuppingerCole Analysts AG

Labelling and Training

The process of labelling data and using it for training



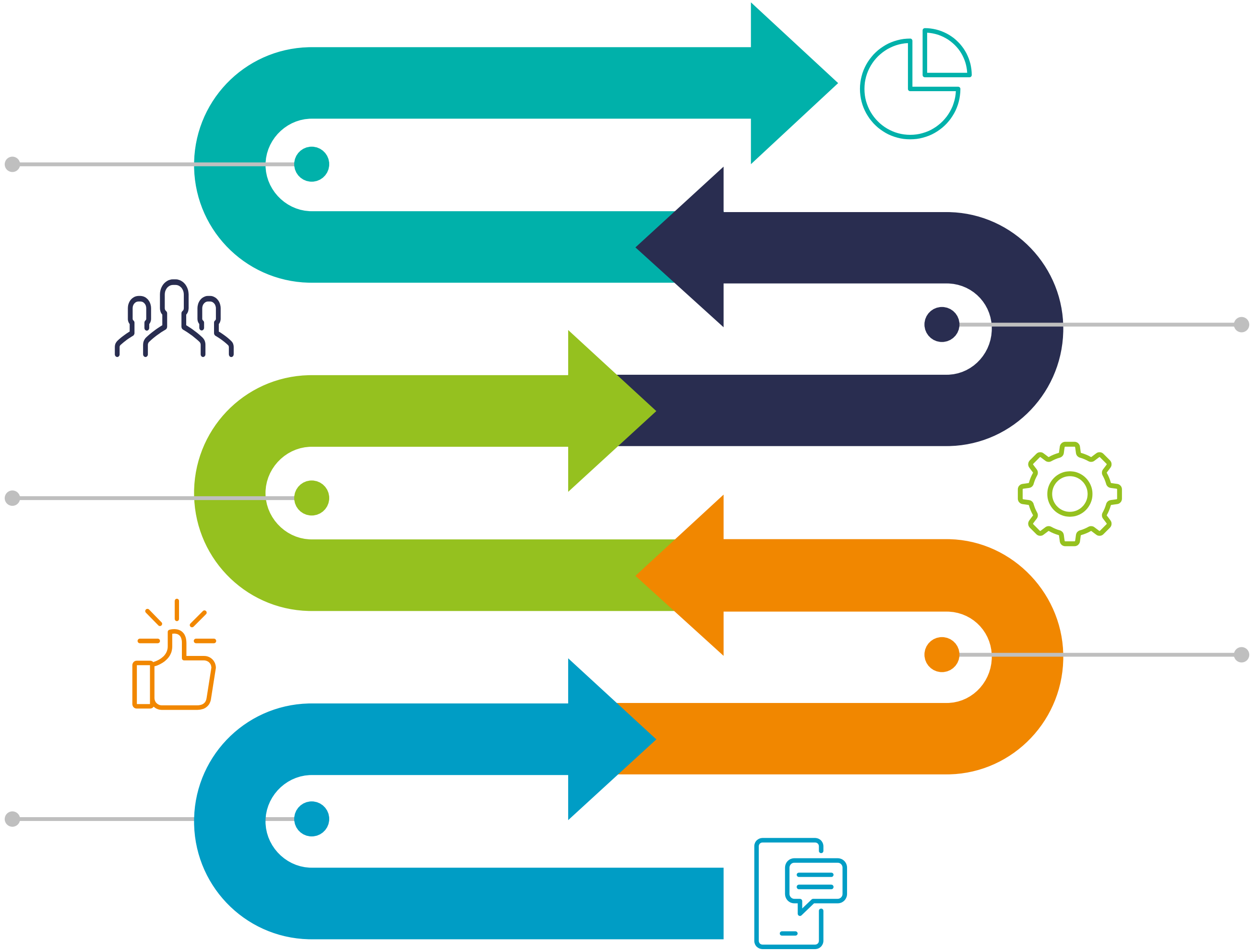
AI Governance Challenges

An overview of the major challenges

- ⊗ Explainability.
- ⊗ Data Privacy.
- ⊗ Data - Bias.
- ⊗ Lifecycle management
- ⊗ Culture and Ethics
- ⊗ Human Involvement
- ⊗ Adversarial Attack
- ⊗ Internal Risk Management

Explainability

Models can say what but not why – LIME Local Interpretable Model-Agnostic Explanations



For Image Classification

LIME finds the minimal number of pixels that achieves the highest probability for the given explanation .

Explaining a black box model

A black box model can be explained retrospectively with additional algorithms.

Training data

The training data must first be understandable in a real-world context.

Explainable AI

Local Explanation

A model that is applied to interpret an individual decision, like LIME or feature selection.

Interpretable or black box?

Some ML models (like decision trees) are interpretable. Others give no indication how or why a certain decision was made.

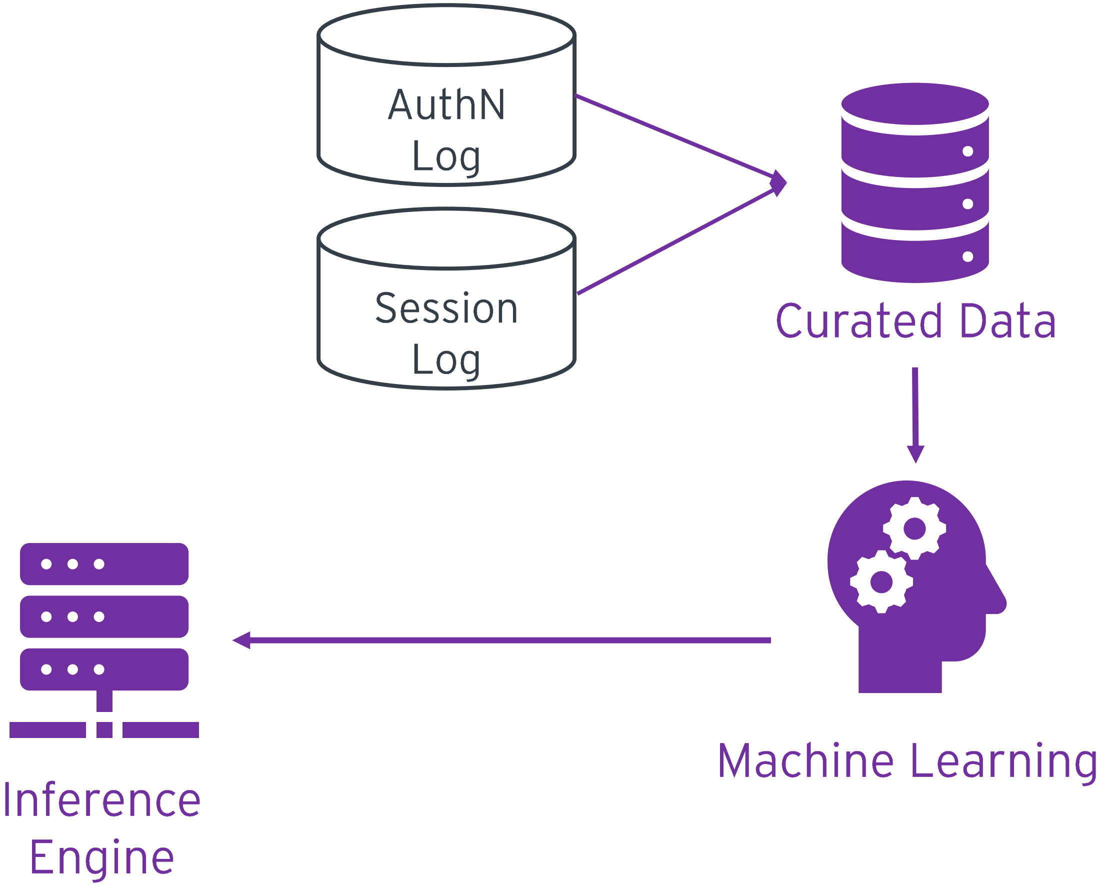
Data Privacy (GDPR Example)

Is it lawful to use the data for this purpose?

Test (1) Is the data "Personal Data"? **YES**

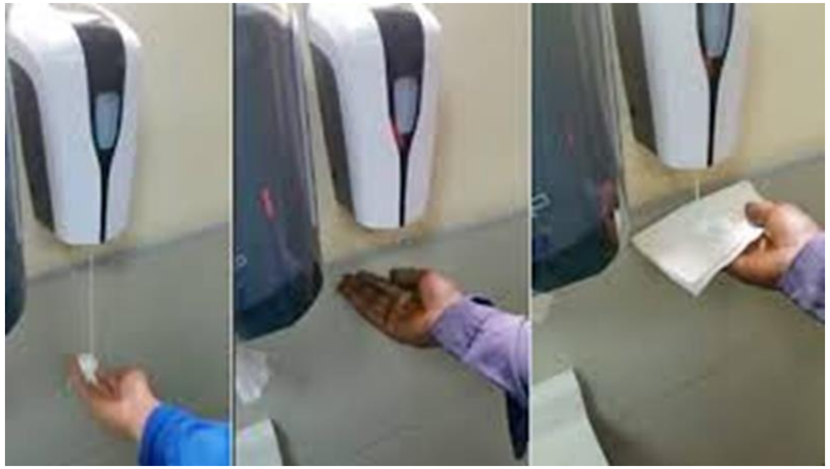
Test (2) is this "Processing" under GDPR? **YES**

Test (3) is this processing fair and lawful under GDPR? **MAYBE**

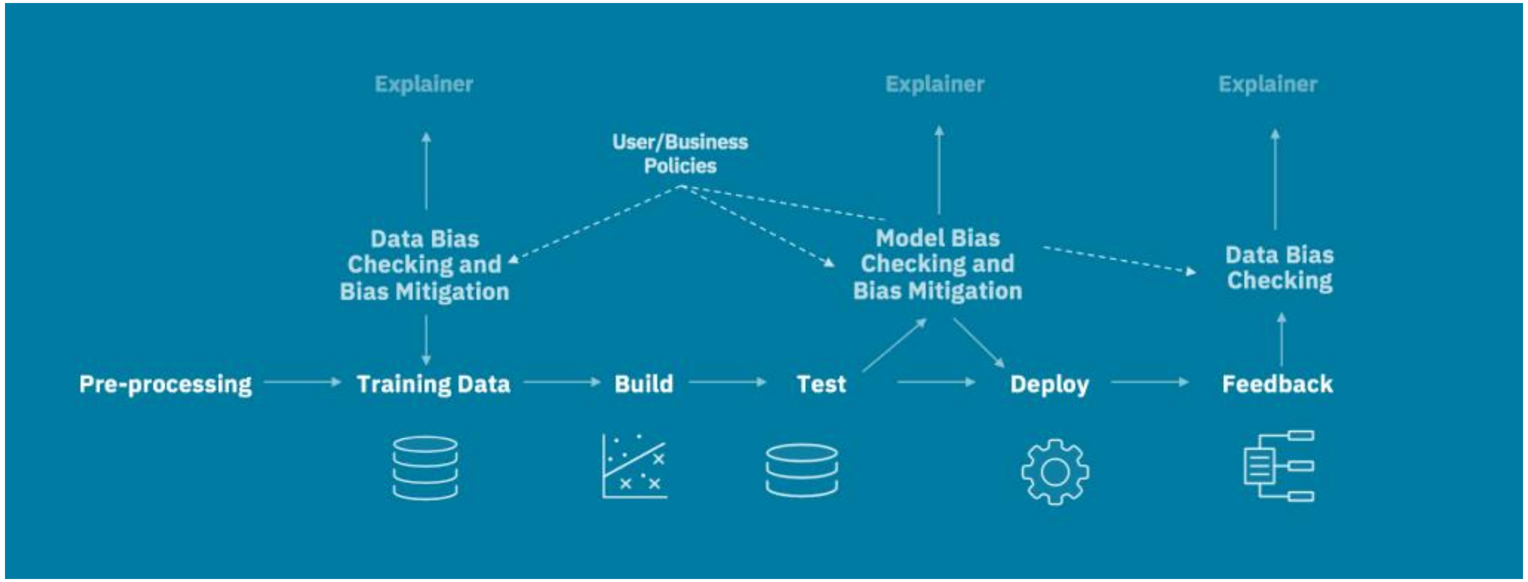


Mitigating Bias

How do you enable Fairness / Accountability / Transparency



2017, sensor failing to detect a dark hand

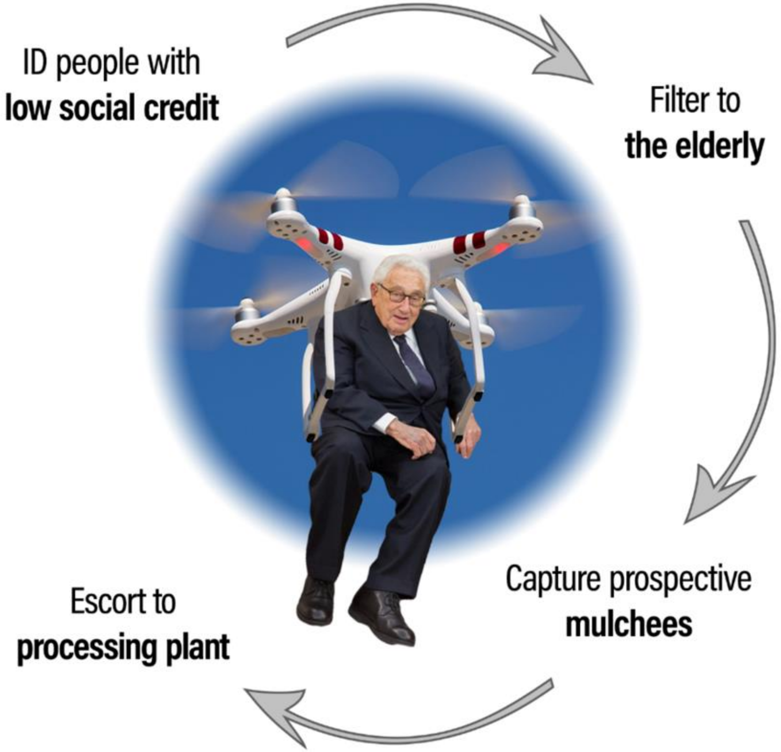


https://aif360.mybluemix.net/

Is the FAT Approach Sufficient?

An example misusing the FAT rules

Fairness / Accountability / Transparency - does not guarantee an ethical AI system!



ID people with **low social credit**

Filter to **the elderly**

Capture prospective **mulchees**

Escort to **processing plant**

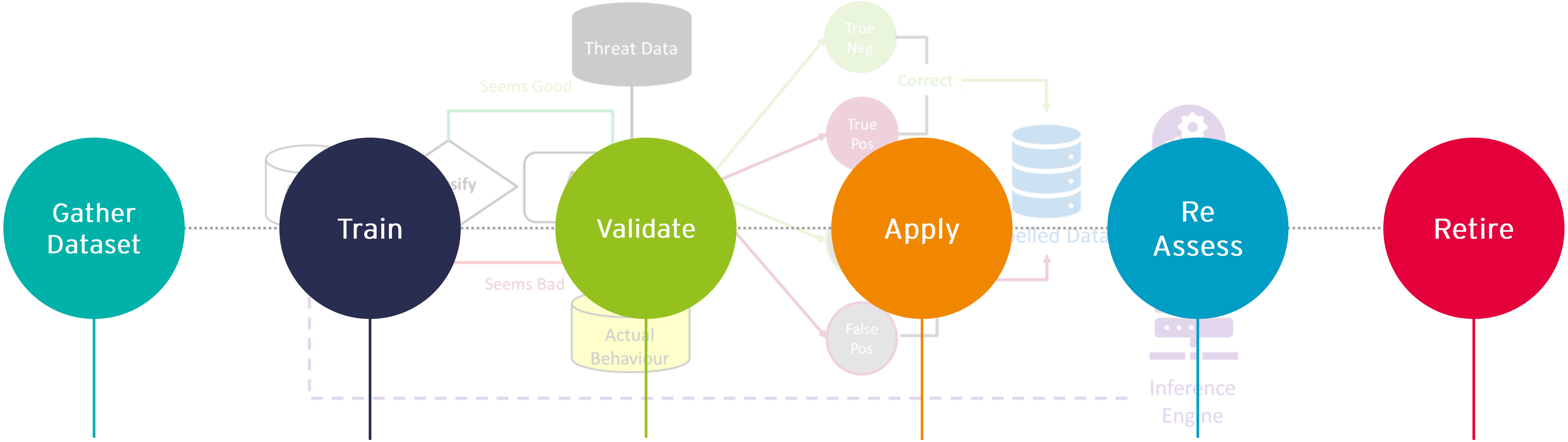
Logan-Nolan Industries

Helping Humanity Make Ends Meat

A Mulching Proposal: Analysing and Improving an Algorithmic System for Turning the Elderly into High-Nutrient Slurry

Lifecycle Management

It is a journey not a destination - ML models learn, change, and drift over time



Gather Dataset

Train

Validate

Apply

Re Assess

Retire

Gather Datasets
Ensure suitability.

Training
Representative dataset.

Validation
Confirm suitability with unseen dataset.

Implementation
Use real-time data for operational purpose.

Assessment
Model should undergo periodic assessment.

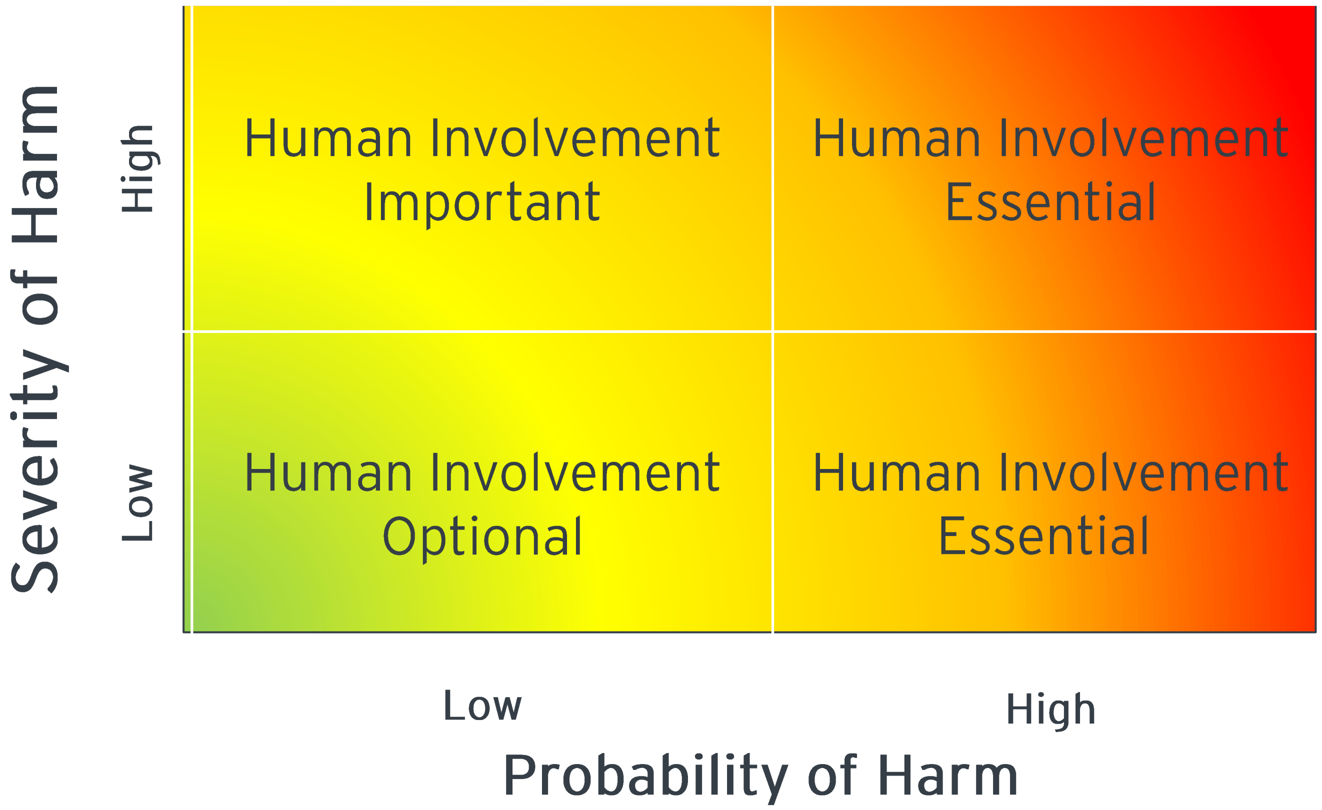
Retirement
Model and all data should be safely disposed of.

Human involvement in AI decision-making

What level of human involvement should there be?

Physicians of the Utmost Fame
Were called at once; but when they came
They answered, as they took their Fees,
“There is no cure for this Disease.
Henry will very soon be dead.”

- Hilaire Belloc



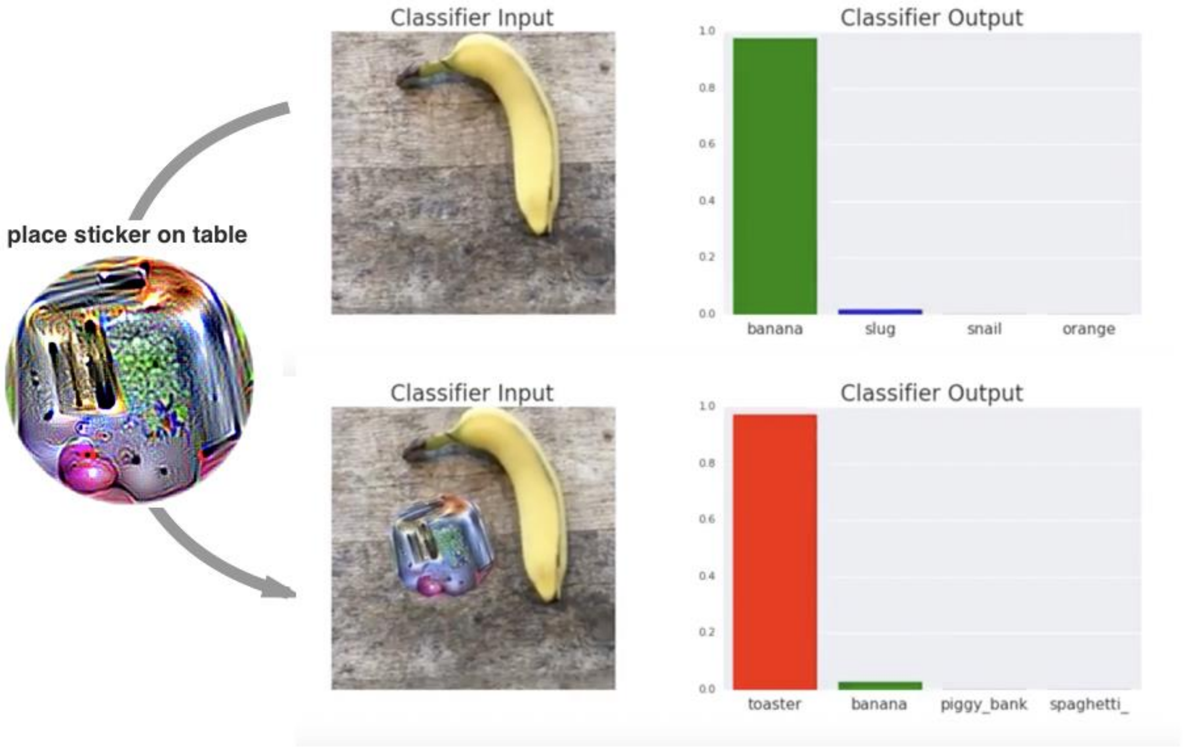
Severity of Harm / Probability of Harm	Low	High
High	Human Involvement Important	Human Involvement Essential
Low	Human Involvement Optional	Human Involvement Essential

From the Singapore Model AI Governance Framework

Adversarial Attack

How easy is it to deliberately confuse the ML system?

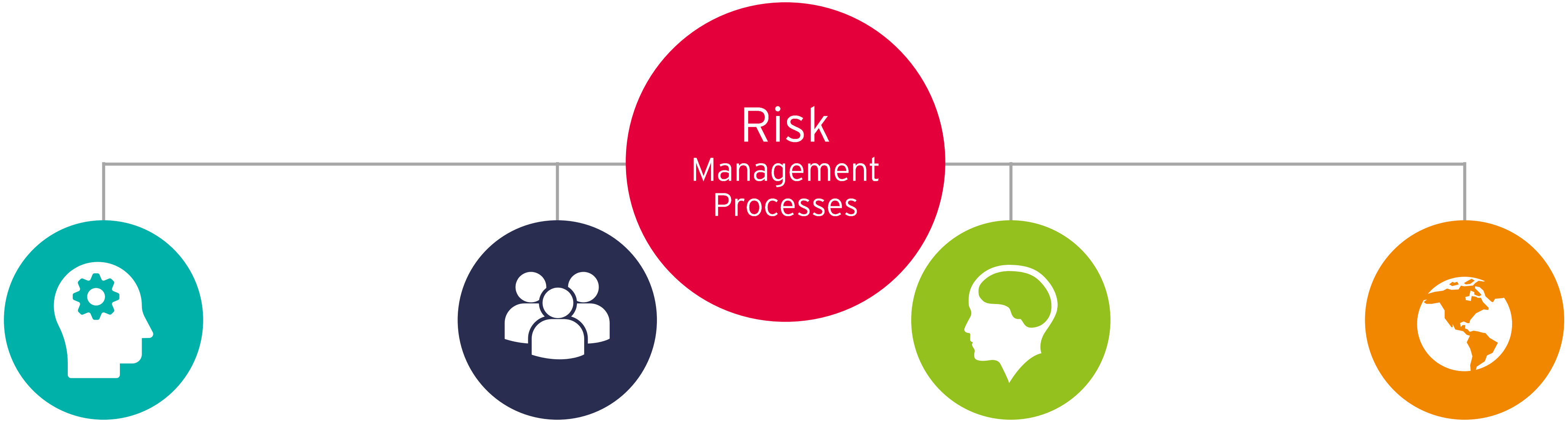
- Neural networks do not depend upon understanding
- They simply look for a match with a pattern
- Changes easily detected by a person can confuse ML
- This can be used against ML.



A survey of practical adversarial example attacks

Internal Risk Management

How to manage the risks of AI to the organization?



Risk Management Processes

Multi-level Assurance Framework

Managed by CSO & DPO, leads department head in formulating and overseeing comprehensive controls.

Development Process

Combined effort from engineering, product, sales, research teams. Follows a "concept, consult, approve" format.

Approval Board

Combined effort from engineering, product, sales, research teams. Follows a "concept, consult, approve" format. Comprised of CTO, CSO, CIO and representatives from each department. Approves AI models.

Data Provenance Controls

Continually review data points to ensure alignment with model's business purpose, check validity of datasets, and manage lifecycle aspects.

Agenda

01 | What is AI

02 | AI for governance

03 | Governance of AI - Example Scenario

04 | AI Governance Frameworks

Comparison of Global AI Frameworks

Alignment of Non-Binding Standards

Frameworks	Human Involvement	Agile Involvement	Ethical	Technical Robustness	Data Privacy	Accounta-bility	Legal Legitimacy	Social/ Environ-mental	Human-Centric	Internal Risk Mngmt
Layered Model for AI Governance (2017, Harvard)			■	■			■	■		
China's New Generation AI Governance Principles (2019)		■	■	■	■	■	■	■	■	
OECD AI Principles (2019)		■	■	■		■	■	■		
EU Guidelines on Ethics in AI (2020)	■		■	■	■	■	■	■	■	
Google Perspectives on Issues in AI Governance (2020)		■	■	■						■
Singapore Model Framework (2020)		■	■	■		■		■	■	■

Imagining Future Regulations on AI – Similarities to the GDPR? - KuppingerCole

Singapore Model AI Governance Framework

Guiding Principles:

- Decisions should be Explainable, Transparent, Fair
- AI Systems should be Human-Centric

Internal Governance Structures and Measures

Clear roles and responsibilities, SOPs to monitor and manage risks, staff training

Determining Level of Human-Involvement

Determine the appropriate level of human involvement, minimize risk of harm to users

Operations Management

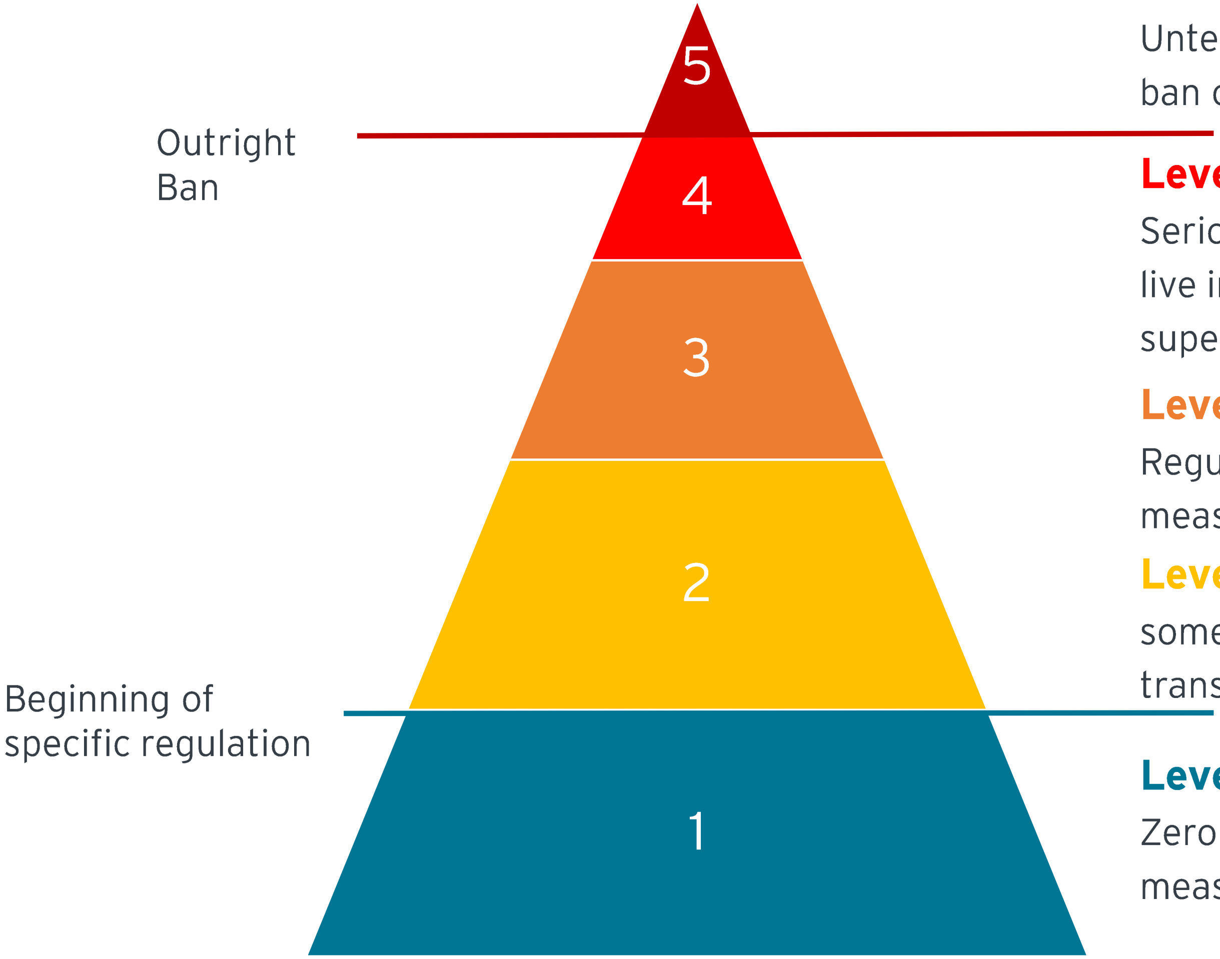
Minimize bias in data and model, risk-based approach to explainability, robustness, lifecycle management

Stakeholder Interaction and Communication

Make AI policies known to users, allow user feedback, make communications human-readable

5 Levels of Risk

Adapted from the Opinion of the Data Ethics Commission, European Union



Level 5

Untenable potential for harm, complete or partial ban of an algorithmic system

Level 4

Serious potential for harm, additional measures (e.g. live interface for "always on" oversight by supervisory institutions)

Level 3

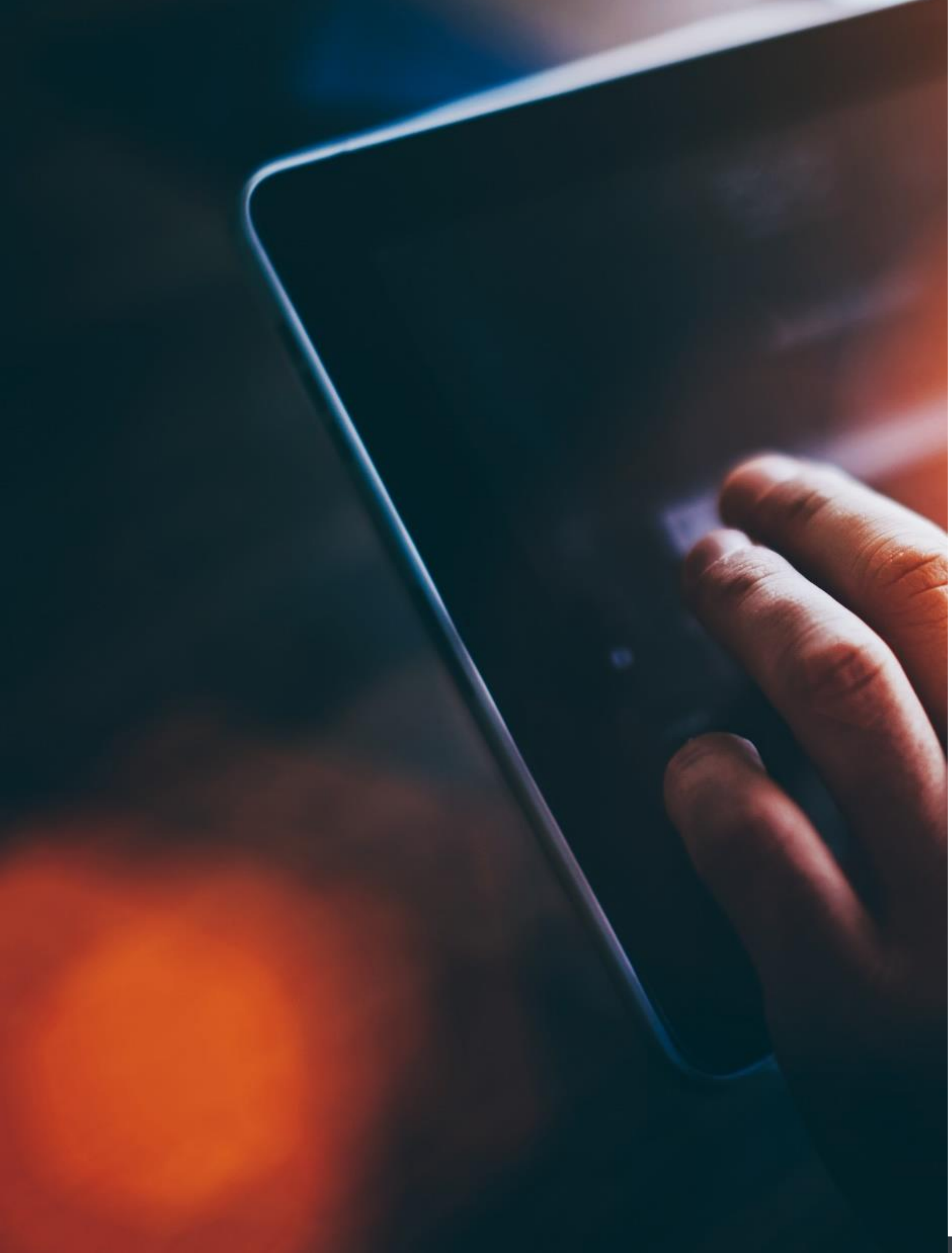
Regular or significant potential for harm, additional measures (e.g. ex-ante approval procedures)

Level 2

some potential for harm, some measures (e.g. transparency obligations, audit procedures)

Level 1

Zero or negligible potential for harm, no special measures



Summary

AI is useful but also needs careful governance

01 AI is not new, but technology has made it more practical

02 AI introduces new governance challenges

03 Some governance frameworks exist as a precursor to legislation

04 Choose how you apply AI with great care

KuppingerCole Analysts AG
Wilhelmstr. 20 - 22
65185 Wiesbaden | GERMANY

P: +49 | 211 - 23 70 77 - 0
F: +49 | 211 - 23 70 77 – 11

E: info@kuppingercole.com
www.kuppingercole.com