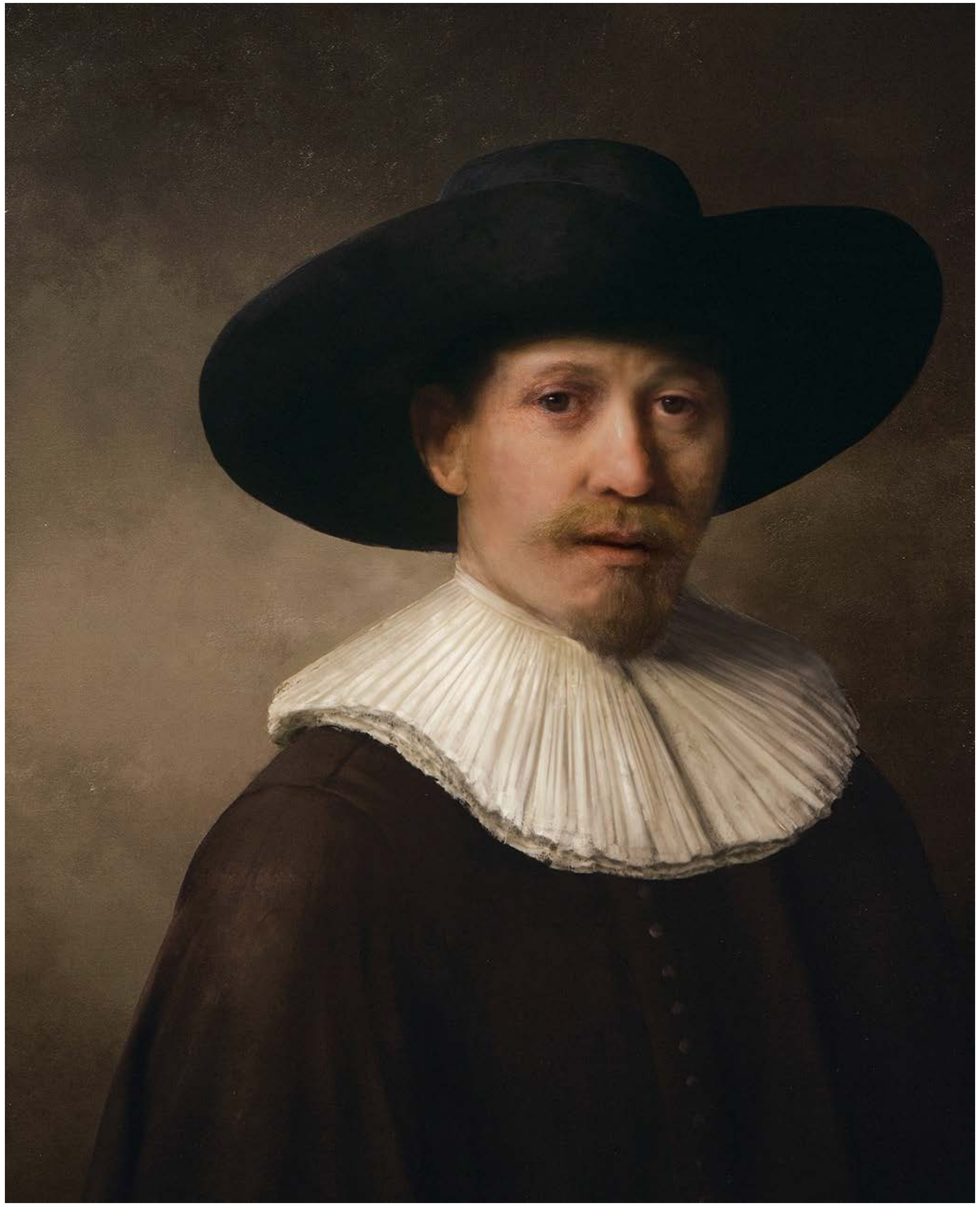
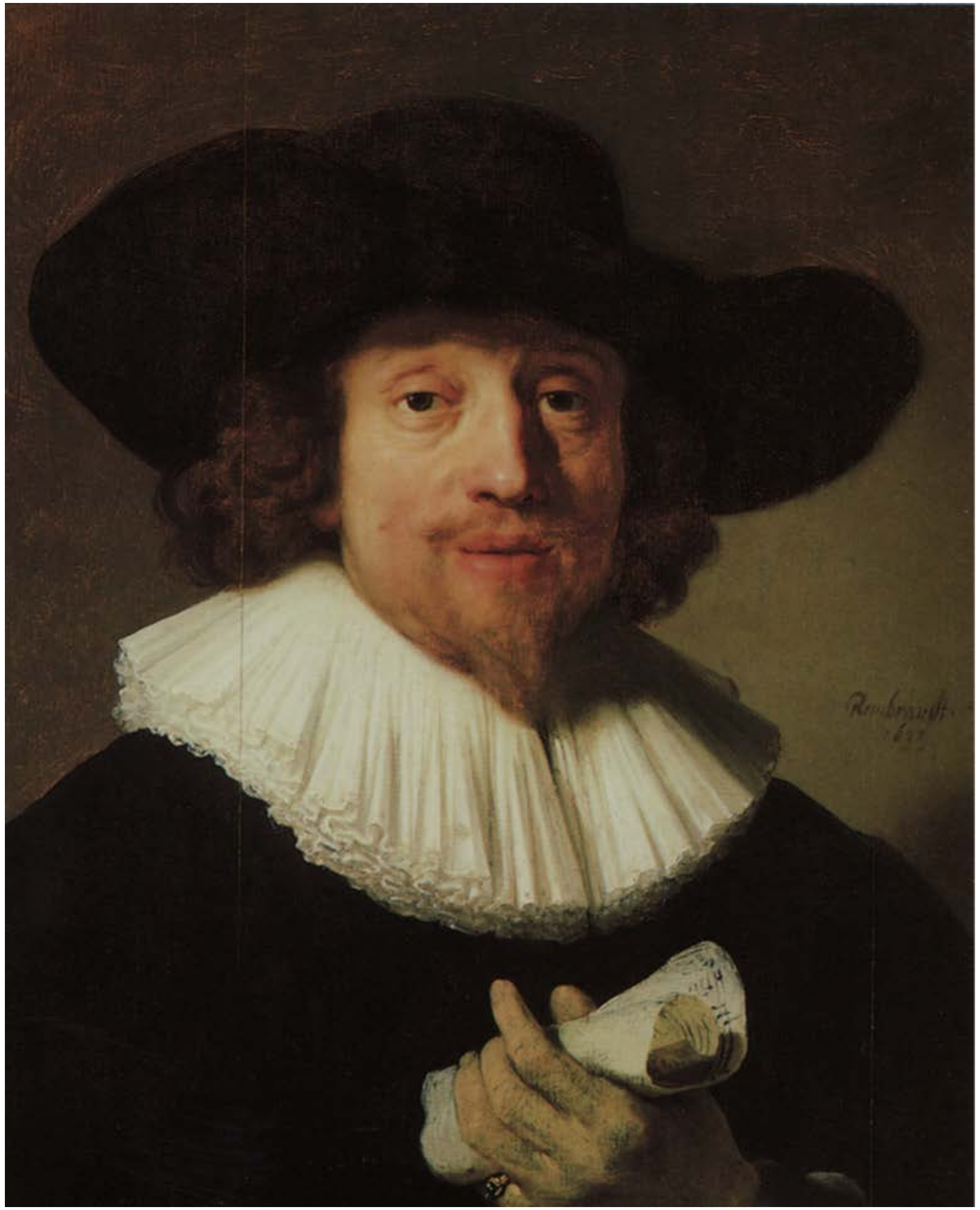
Dialogveranstaltung KI im Mittelstand – Open Session

EU AI ACT und Normen: Schlüssel zur einer vertrauenswürdigen KI

Dr. Andreas Hauser (AIQURIS) & Dr. Wolfgang Hildesheim (IBM)





Rembrandt.

Ghost Musician Records

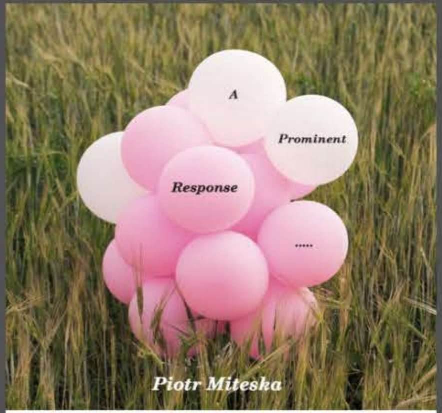
Music your ears won't believe is true.

DIN

Releases

Charles Bolt

Ana Olgica



Piotr Miteska

Sigimund

They Dream By Day

Charlie Key

1. The European AI Act

Goal is the acceleration of innovation & the mitigation of risk

Europäische KI Verordnung (“EU AI Act”) in 2024

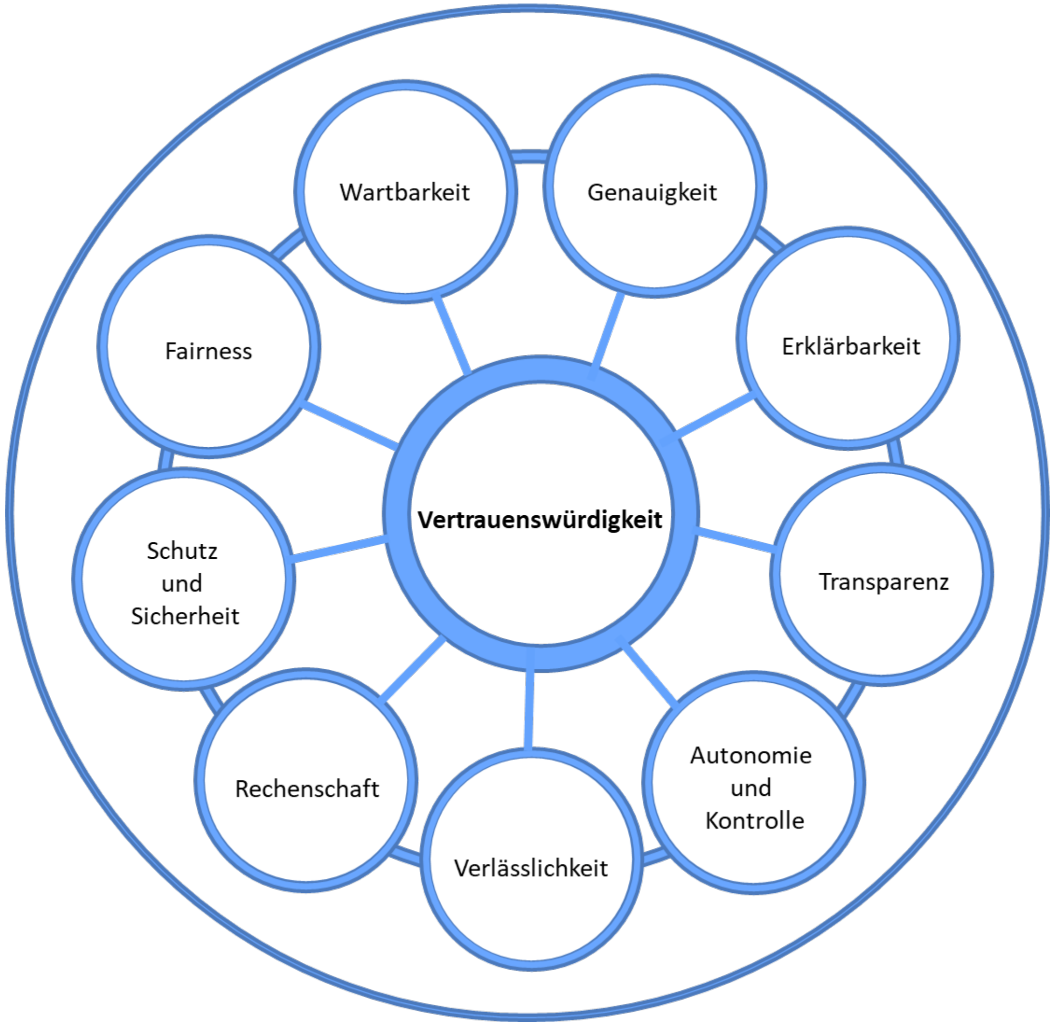
Regulierung auf der Basis von Risikoklassen

Vertrauenswürdigkeit von KI wird der Wettbewerbsvorteil der EU werden

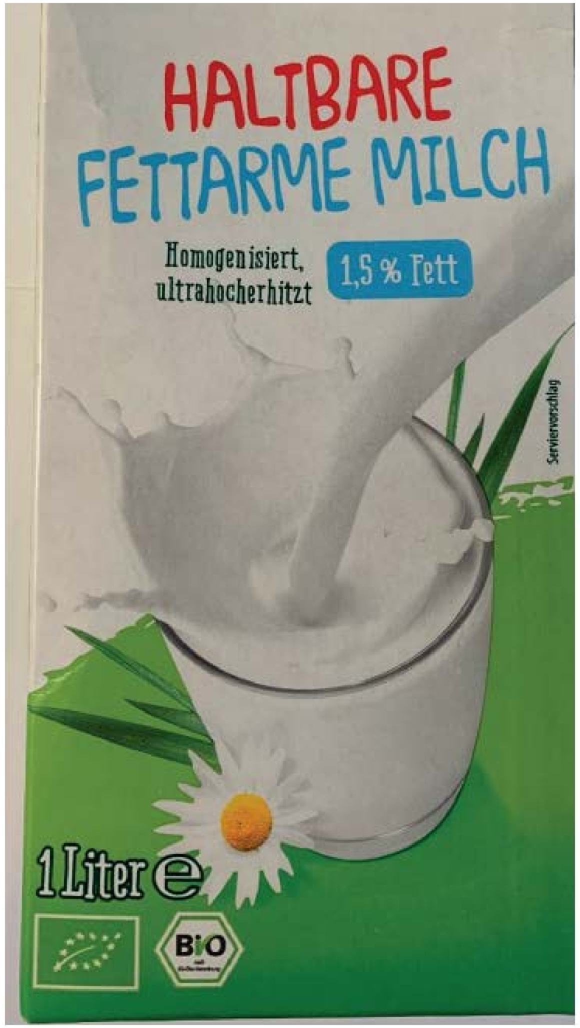
Risikoklasse gemäß AI Act

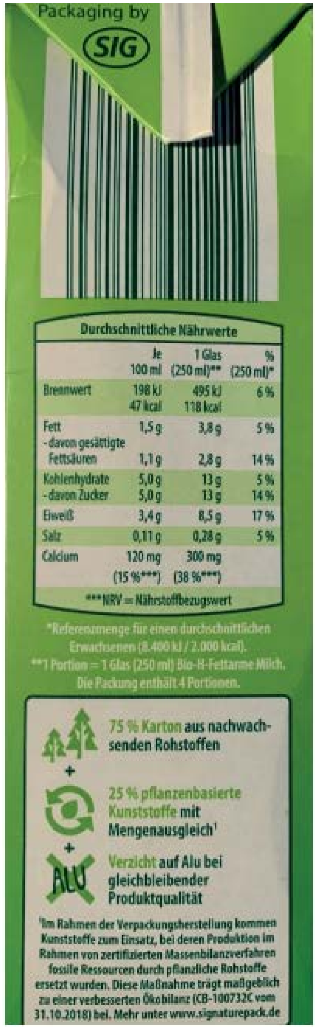
- **Verbotene KI-Anwendungen**
- **Hochrisiko-KI-Systeme**
- **KI-Systeme mit bes. Transparenzpflichten**
- **Sonstige**

Erfolgsfaktor Testen & Zertifizieren



Food labelling as an example how to create trust





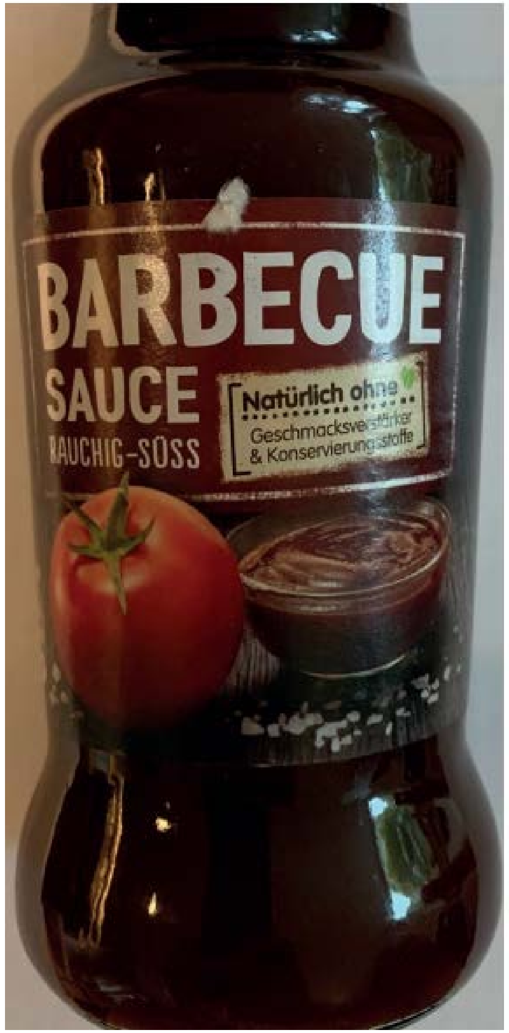
Durchschnittliche Nährwerte	Je 100 ml	1 Glas (250 ml)**	% (250 ml)*
Brennwert	198 kJ 47 kcal	495 kJ 118 kcal	6 %
Fett	1,5 g	3,8 g	5 %
- davon gesättigte Fettsäuren	1,1 g	2,8 g	14 %
Kohlenhydrate	5,0 g	13 g	5 %
- davon Zucker	5,0 g	13 g	14 %
Eiweiß	3,4 g	8,5 g	17 %
Salz	0,11 g	0,28 g	5 %
Calcium	120 mg (15 %***)	300 mg (38 %***)	

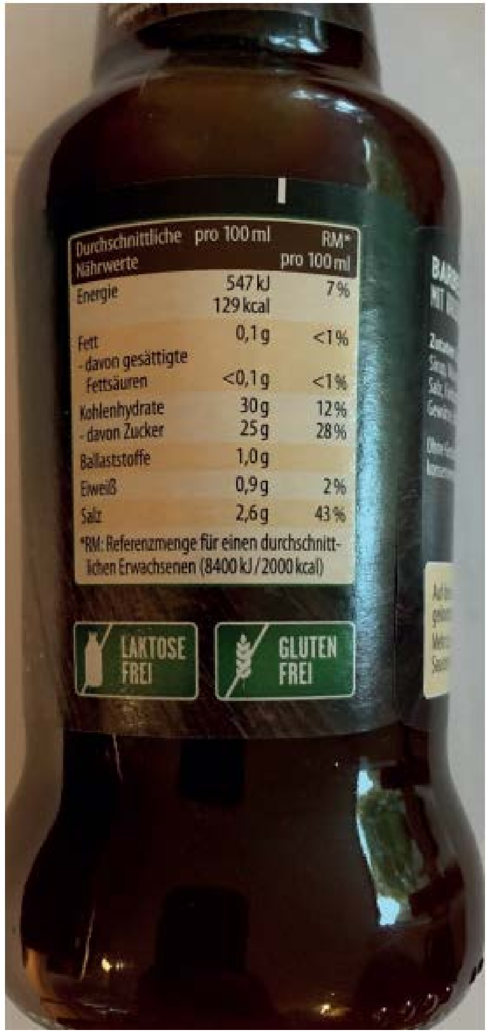
***NRV = Nährstoffbezugswert

*Referenzmenge für einen durchschnittlichen Erwachsenen (8.400 kJ / 2.000 kcal).
**1 Portion = 1 Glas (250 ml) Bio-H-Fettarme Milch. Die Packung enthält 4 Portionen.

75 % Karton aus nachwachsenden Rohstoffen
+
25 % pflanzenbasierte Kunststoffe mit Mengenausgleich[1]
+
ALU
Verzicht auf Alu bei gleichbleibender Produktqualität

[1] Im Rahmen der Verpackungsherstellung kommen Kunststoffe zum Einsatz, bei deren Produktion im Rahmen von zertifizierten Massenbilanzverfahren fossile Ressourcen durch pflanzliche Rohstoffe ersetzt wurden. Diese Maßnahme trägt maßgeblich zu einer verbesserten Ökobilanz (CB-100732C vom 31.10.2018) bei. Mehr unter www.signaturepack.de

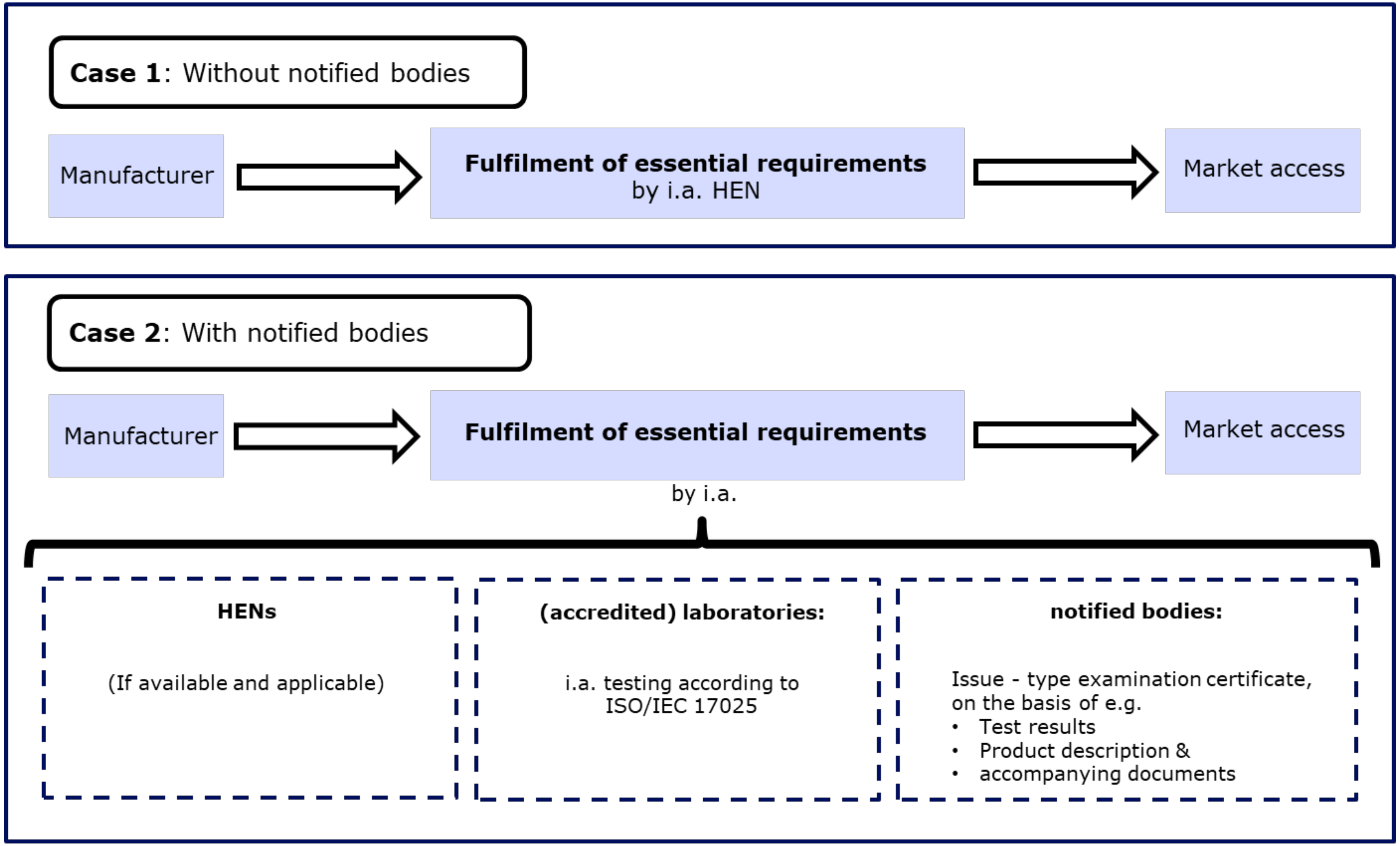




Durchschnittliche Nährwerte	pro 100 ml	RM* pro 100 ml
Energie	547 kJ 129 kcal	7%
Fett	0,1 g	<1%
- davon gesättigte Fettsäuren	<0,1 g	<1%
Kohlenhydrate	30 g	12 %
- davon Zucker	25 g	28 %
Ballaststoffe	1,0 g	
Eiweiß	0,9 g	2 %
Salz	2,6 g	43 %

*RM: Referenzmenge für einen durchschnittlichen Erwachsenen (8400 kJ / 2000 kcal)

Conformity assessment (1/2)
Without and with notified bodies



Case 1: Without notified bodies

Manufacturer

Fulfilment of essential requirements
by i.a. HEN

Market access

Case 2: With notified bodies

Manufacturer

Fulfilment of essential requirements

Market access

by i.a.

HENs

(If available and applicable)

(accredited) laboratories:

i.a. testing according to
ISO/IEC 17025

notified bodies:

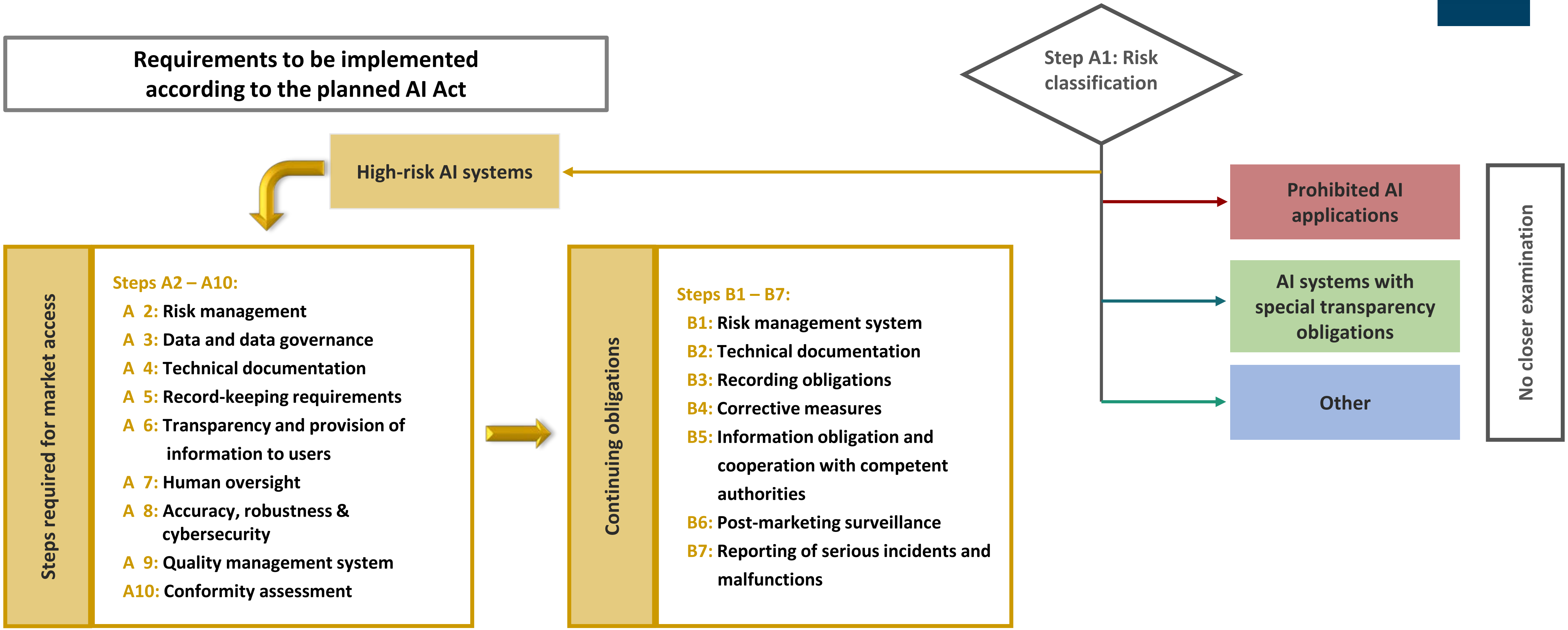
Issue - type examination certificate, on the basis of e.g.

- Test results
- Product description &
- accompanying documents

Figure 16.2: Conformity assessment without and with NoBos.

Conformity assessment and market surveillance

"step-by-step" requirements


Requirements to be implemented according to the planned AI Act
Step A1: Risk classification
High-risk AI systems
Prohibited AI applications
AI systems with special transparency obligations
Other
No closer examination
Steps required for market access
Continuing obligations

Steps A2 – A10:

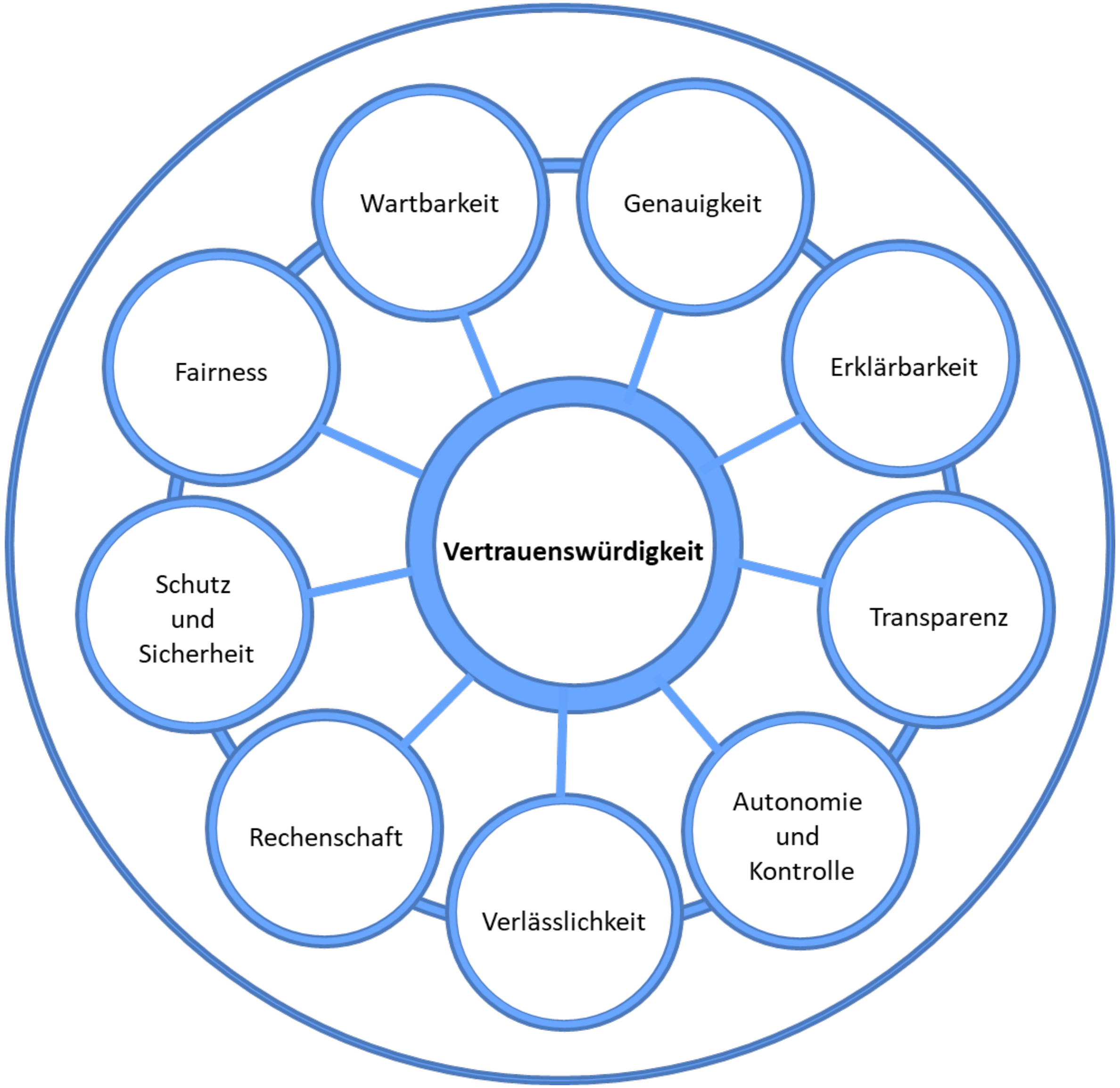
A 2: Risk management
A 3: Data and data governance
A 4: Technical documentation
A 5: Record-keeping requirements
A 6: Transparency and provision of information to users
A 7: Human oversight
A 8: Accuracy, robustness & cybersecurity
A 9: Quality management system
A10: Conformity assessment

Steps B1 – B7:

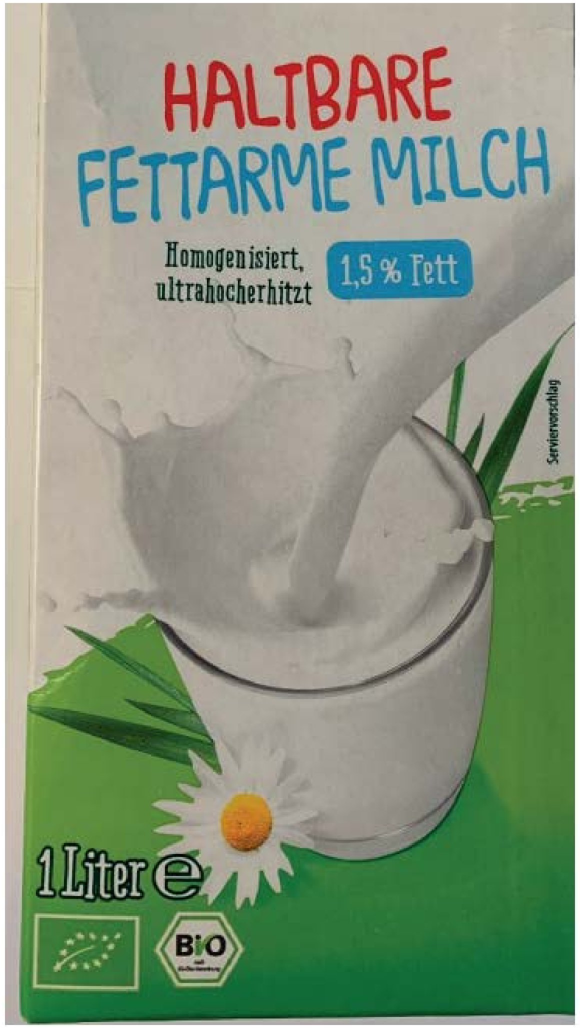
B1: Risk management system
B2: Technical documentation
B3: Recording obligations
B4: Corrective measures
B5: Information obligation and cooperation with competent authorities
B6: Post-marketing surveillance
B7: Reporting of serious incidents and malfunctions

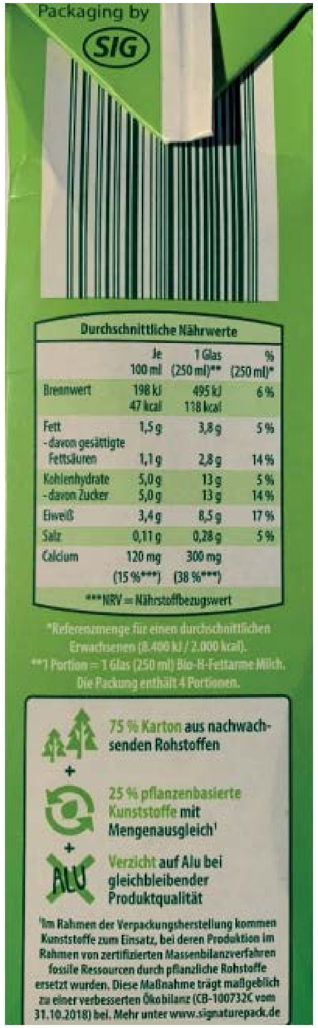
Figure 17.1: Step-to-step guide for implementing the requirements of the planned AI Act according to [1].

Vertrauenswürdigkeit von KI ist erfolgsentscheidend



Food labelling as an example how to create trust





Durchschnittliche Nährwerte

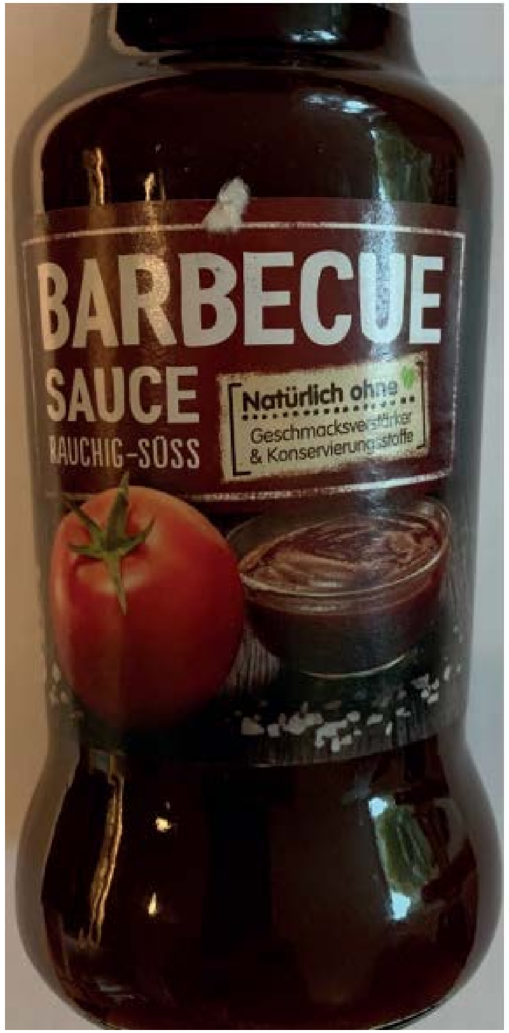
	Je 100 ml	1 Glas (250 ml)**	% (250 ml)*
Brennwert	198 kJ 47 kcal	495 kJ 118 kcal	6 %
Fett	1,5 g	3,8 g	5 %
- davon gesättigte Fettsäuren	1,1 g	2,8 g	14 %
Kohlenhydrate	5,0 g	13 g	5 %
- davon Zucker	5,0 g	13 g	14 %
Eiweiß	3,4 g	8,5 g	17 %
Salz	0,11 g	0,28 g	5 %
Calcium	120 mg (15 %***)	300 mg (38 %***)	

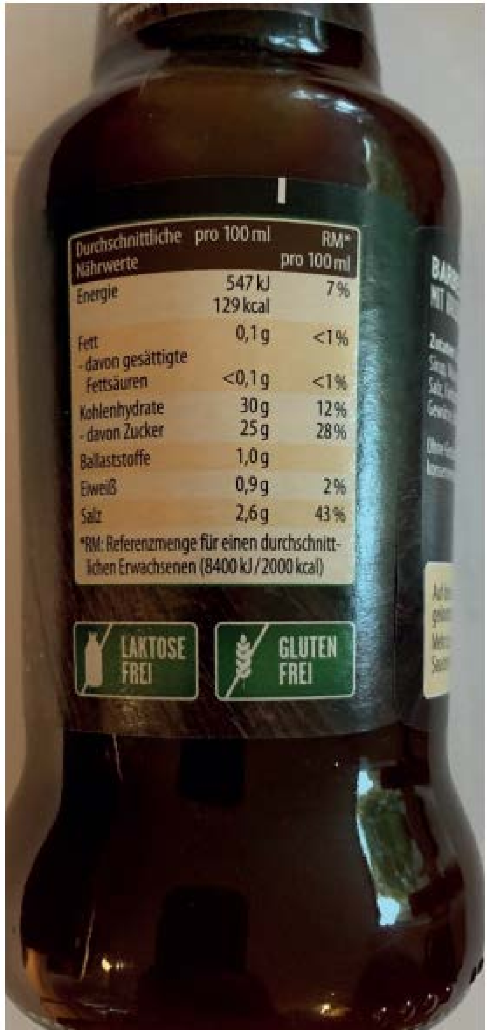
***NRV = Nährstoffbezugswert

*Referenzmenge für einen durchschnittlichen Erwachsenen (8.400 kJ / 2.000 kcal).
**1 Portion = 1 Glas (250 ml) Bio-H-Fettarme Milch. Die Packung enthält 4 Portionen.

75 % Karton aus nachwachsenden Rohstoffen
+
25 % pflanzenbasierte Kunststoffe mit Mengenausgleich[1]
+
ALU
Verzicht auf Alu bei gleichbleibender Produktqualität

[1] Im Rahmen der Verpackungsherstellung kommen Kunststoffe zum Einsatz, bei deren Produktion im Rahmen von zertifizierten Massenbilanzverfahren fossile Ressourcen durch pflanzliche Rohstoffe ersetzt wurden. Diese Maßnahme trägt maßgeblich zu einer verbesserten Ökobilanz (CB-100732C vom 31.10.2018) bei. Mehr unter www.signaturepack.de





Durchschnittliche Nährwerte	pro 100 ml	RM* pro 100 ml
Energie	547 kJ 129 kcal	7%
Fett	0,1 g	<1%
- davon gesättigte Fettsäuren	<0,1 g	<1%
Kohlenhydrate	30 g	12 %
- davon Zucker	25 g	28 %
Ballaststoffe	1,0 g	
Eiweiß	0,9 g	2 %
Salz	2,6 g	43 %

*RM: Referenzmenge für einen durchschnittlichen Erwachsenen (8400 kJ / 2000 kcal)

Aktueller Kenntnisstand und Nutzung von KI

Kenntnisstand zum AI Act

- Hoch
- Mittel
- Gering

Reifegrad KI Anwendungen

- Hoch (im produktiven Betrieb)
- Mittel (in Vorbereitung)
- Gering/keine

→ In welchem Bereich/Branche?

Kompatibilität/Konformität Ihrer Prozesse mit dem AI Act

- Ja
- Nein
- unsicher

→ Weil, ggfs. Kommentar

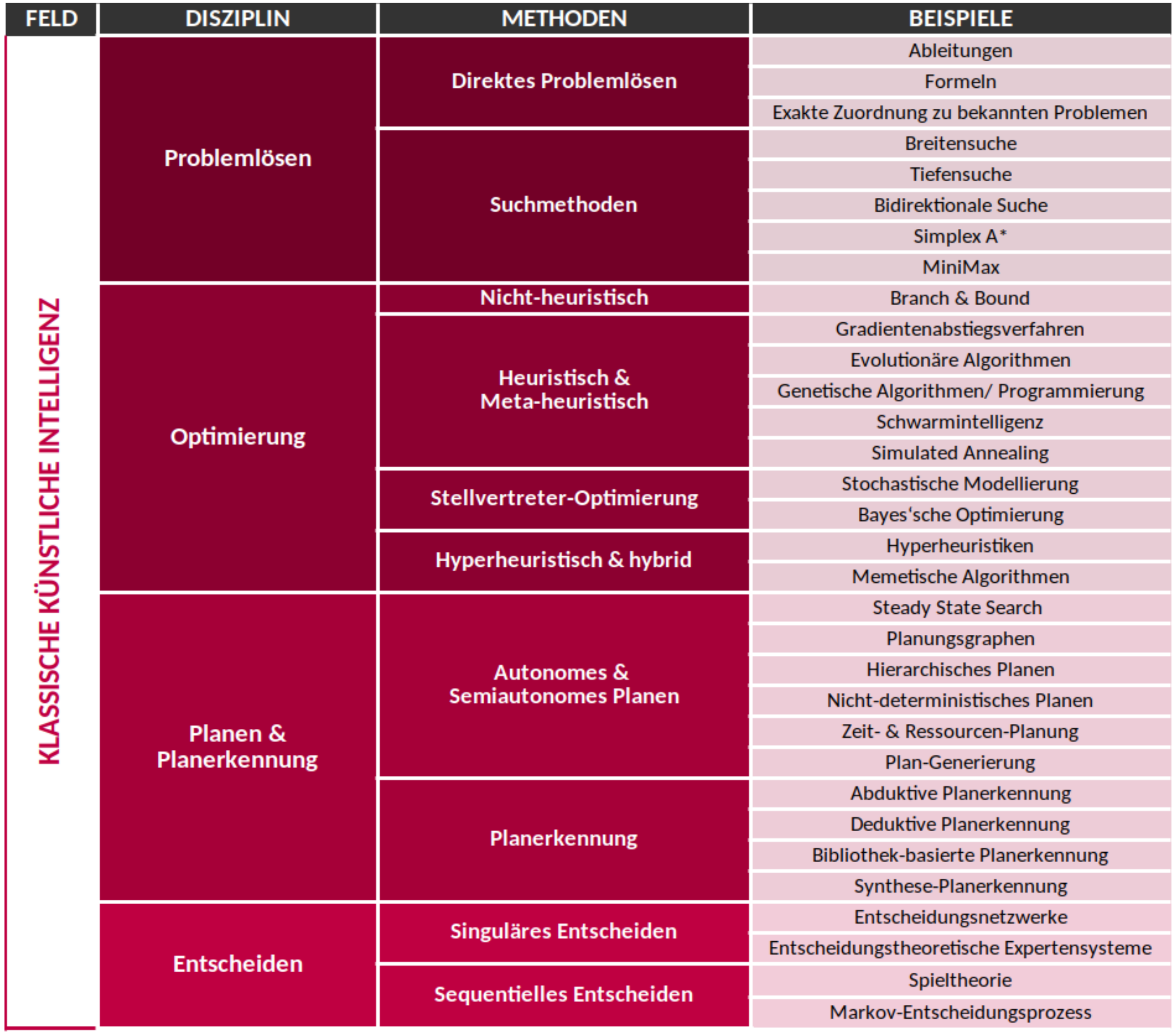
2. AI labels, transparency & norms as competitive advantage

Example 1 - AI solution

AI in shipping (Fraunhofer CML)

Figures 2-5: Typical tasks and roles for the crew composition on a ship.

Methodenspektrum: Klassische KI



FELD	DISZIPLIN	METHODEN	BEISPIELE
KLASSISCHE KÜNSTLICHE INTELLIGENZ	Problemlösen	Direktes Problemlösen	Ableitungen
			Formeln
			Exakte Zuordnung zu bekannten Problemen
		Suchmethoden	Breitensuche
			Tiefensuche
			Bidirektionale Suche
			Simplex A*
			MiniMax
	Optimierung	Nicht-heuristisch	Branch & Bound
		Heuristisch & Meta-heuristisch	Gradientenabstiegsverfahren
			Evolutionäre Algorithmen
			Genetische Algorithmen/ Programmierung
			Schwarmintelligenz
			Simulated Annealing
		Stellvertreter-Optimierung	Stochastische Modellierung
			Bayes'sche Optimierung
		Hyperheuristisch & hybrid	Hyperheuristiken
			Memetische Algorithmen
	Planen & Planerkennung	Autonomes & Semiautonomes Planen	Steady State Search
			Planungsgraphen
			Hierarchisches Planen
			Nicht-deterministisches Planen
			Zeit- & Ressourcen-Planung
			Plan-Generierung
		Planerkennung	Abduktive Planerkennung
			Deduktive Planerkennung
			Bibliothek-basierte Planerkennung
			Synthese-Planerkennung
	Entscheiden	Singuläres Entscheiden	Entscheidungsnetzwerke
			Entscheidungstheoretische Expertensysteme
		Sequentielles Entscheiden	Spieltheorie
			Markov-Entscheidungsprozess

AI-assisted crew planning

Efficient crew planning is crucial for ship management companies aiming to op-timize the operation of their fleets both in terms of time and cost. Compliance with international regulations and effective fleet management necessitate a diverse crew with various roles and experiences. Additionally, factors such as border regu-lations, crew rotation, travel logistics, and required vacation times must be taken into account.

To address these complexities, an AI system is employed to strategically allocate ship crews across different time frames. By leveraging advanced algorithms and machine learning, this AI system streamlines planning processes, ensuring both time as well as financial resources are optimized. The application aims for fostering well-being of crew members and facilitating effective time management for both ship utilization and crew vacations.

Focus on legislations/standards:
STCW convention, IMO: MSC.373(93), MSC.486(103), MSC.487(103);
SOLAS, Oil Polution Act.

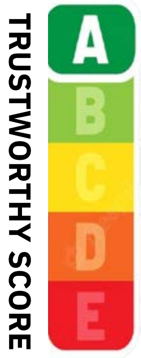
Manufacturer: Fraunhofer Center for Maritime Logistics and Services CML

Risk: Very Low

Application: Improving crew productivity while efficiently meeting crew needs

Capability	Method	Data
	1st step	
Process Knowledge > Factual > Procedural	**Machine Learning** > CNN	**Training Input:** Historic data of ship routes and crew scheduling. **Training Output:** Crew planning model (input for 2nd step) **Model Inference Output**: Rough crew plan
	2nd step	
Process Knowledge > Factual > Procedural	**Traditional AI** > Linear Optimization	**Model Input**: Rough crew plan **Model Output**: Optmized crew plan according to requirements (personal needs, legislation, …)

Data protection provisions:
Confidentiality, integrity and availability of crew, ship and business data are ensured as the application runs on site by the application owner, fulfilling legal requirements.



Example 2 - AI solution

Building Recognition with AI: State Office for Geoinformation and Land Surveying of Lower Saxony (LGLN)

Recognition of buildings and outlines

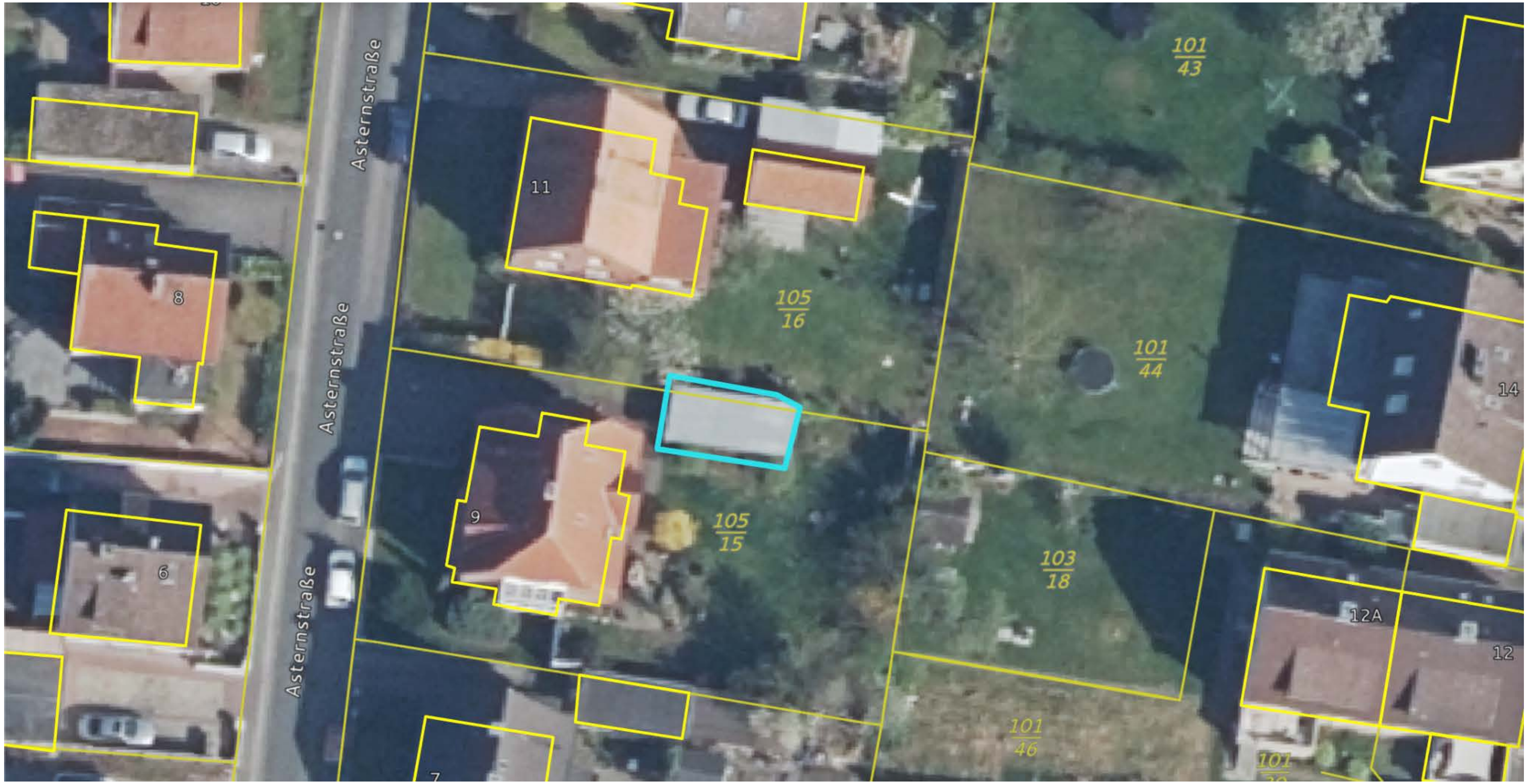


Figure 11.1: Use case for AI: Find the few missing buildings (here a garage in blue) in thebuildings (here garage in blue) in the mass data of the real estate cadastre(buildings and parcels in yellow).

.

Automatic correction of the digital map

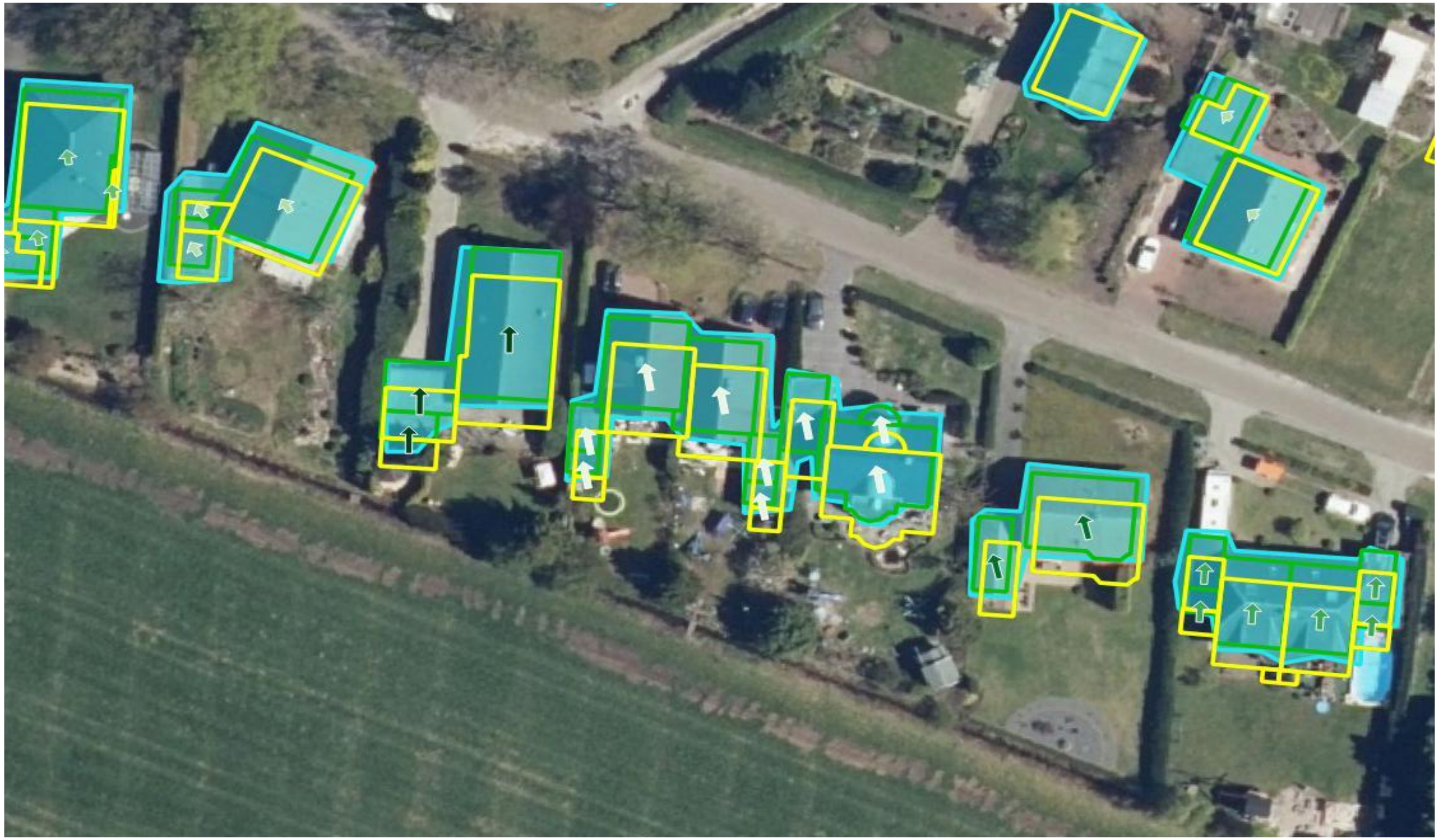
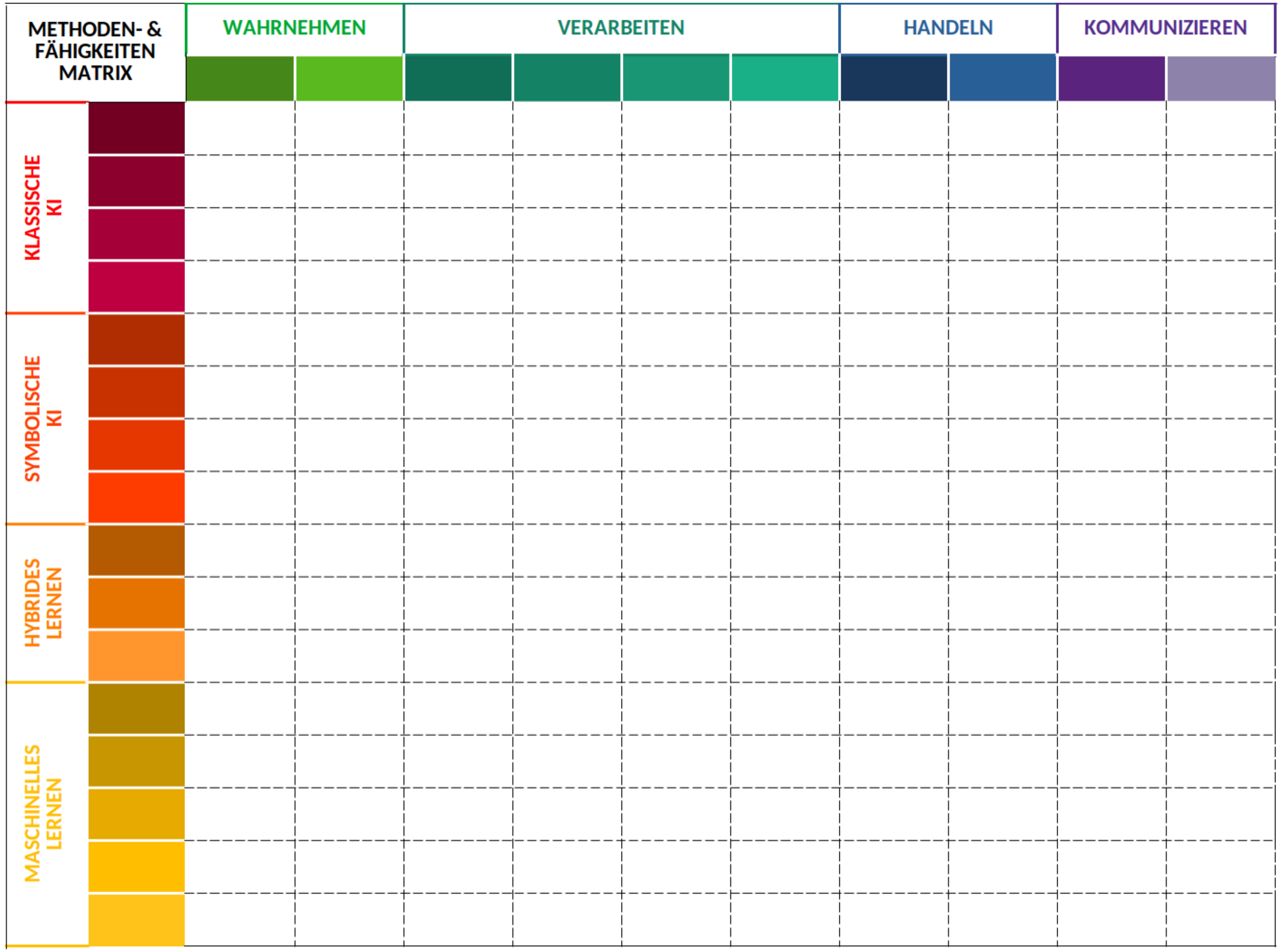


Figure 11.5: Automatic calculation to correct (green circles and arrows) the historical data (yellow) using the arrows) of the historical data (yellow) using object recognition (blue).

KI-Matrix: Methoden-Fähigkeiten



METHODEN- & FÄHIGKEITEN MATRIX		WAHRNEHMEN		VERARBEITEN				HANDELN		KOMMUNIZIEREN	
KLASSISCHE KI											
SYMBOLISCHE KI											
HYBRIDES LERNEN											
MASCHINELLES LERNEN											

Figure 5.3: Two-dimensional representationof AI methods and AI capabilities.

AI Matrix: Methods-Capabilities-Criticality

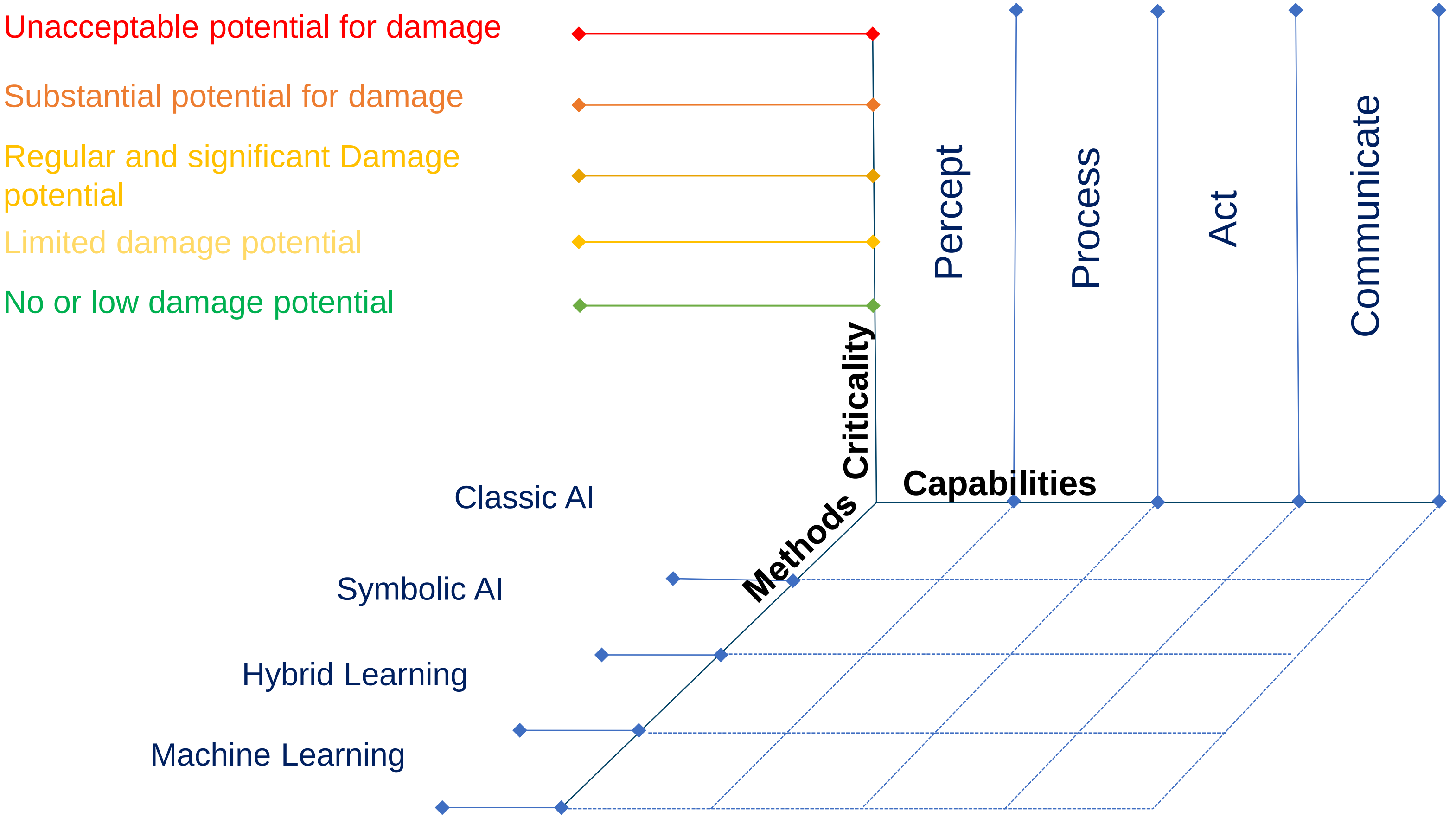


Figure 5.4: Three-dimensional representation of the AI=MC² taxonomy.

DIN

Artificial Intelligence

The AI model recognises objects on satellite images. The model is trained on building locations, building outlines, building types and other building properties.The satellite images are composed of different sensor data: optical images, lidar images.

DIN Standard 4711.4711
CE – AI Label V.0

Manufacturer: LGLN
Model version: V2.3 aus 2022

Risk: Very Low

Application: Object recognition on aerial photographs

Capability	Method	Data
Percept >External >>See	**Machine Learning** >Supervised Learning >>Neural Network	**Training input:** known quality-checked satellite images (optical, infrared) and lidar data Operational data: unknown images
Process >Facts >>Select >>Verify	**Machine Learning** >Supervised Learning >>Neural Network	Building location ,+-xx cm, Building outline ,+- yy cm
Act		
Communicate		

TRUSTWORTHY SCORE
A
B
C
D
E

2 – KI Anwendungen der Teilnehmenden

Was für (geplante) KI Anwendungen sind für Sie relevant?

Bild/Computer Vision

Text/Sprachmodelle

Zahlen/Daten/Vorhersagen

2 – Beispiele für Wettbewerbsvorteile durch AI lables und AI norms

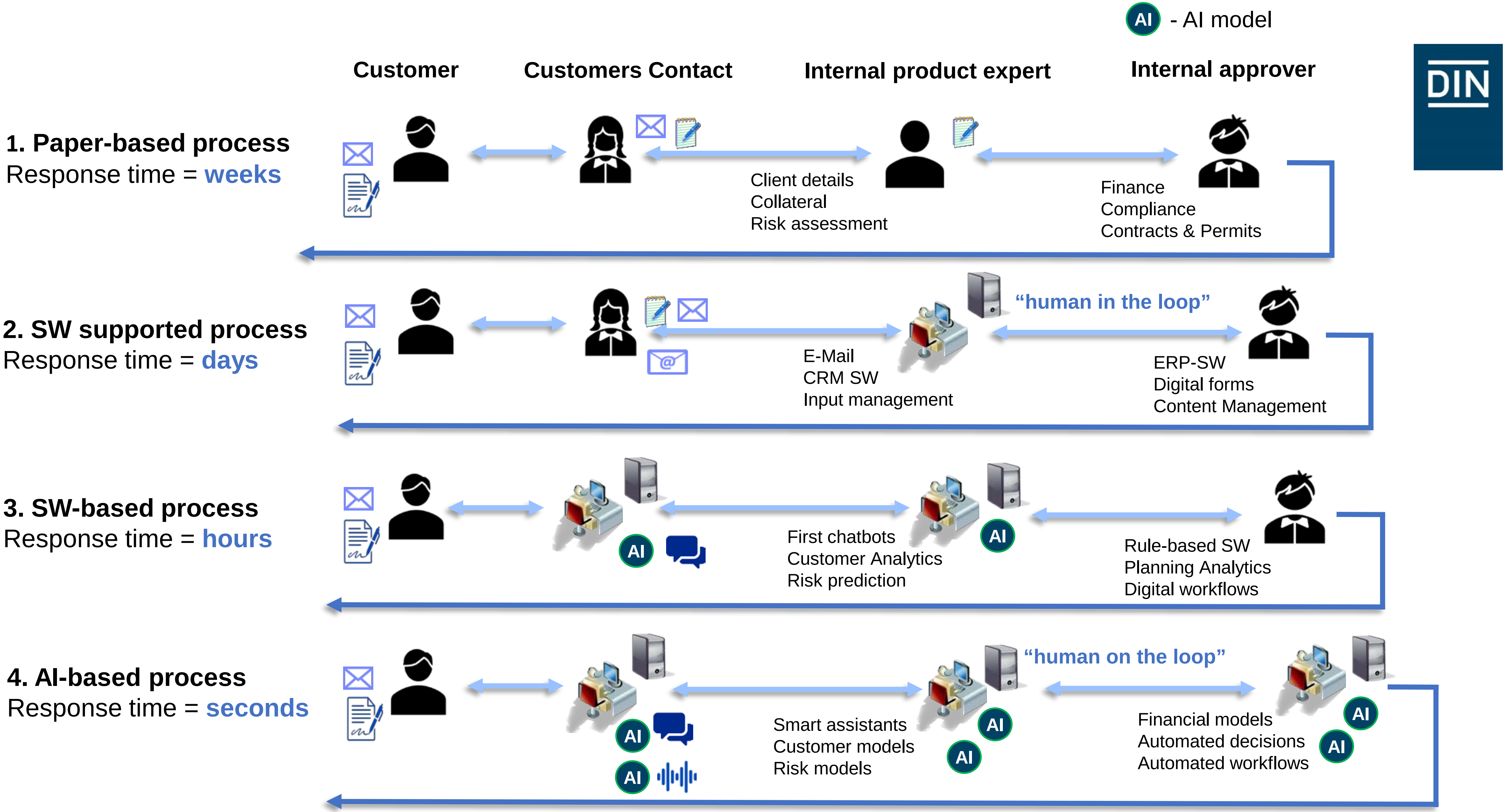
Wie kann das Vertrauen der Kunden in ein Produkt/Service oder dessen Sicherheit zu erhöht werden?

Wie können Sie Vorteile/Wettbewerbsvorteile daraus ziehen?

Welche Herausforderungen können bei der Einführung von KI Labeln und Normen auftreten?

Welche Herausforderungen gab es bei ähnlichen Themen/Labeln?

3. AI Testing as success factor



AI - AI model

Customer | Customers Contact | Internal product expert | Internal approver

1. Paper-based process
Response time = **weeks**

Client details
Collateral
Risk assessment

Finance
Compliance
Contracts & Permits

2. SW supported process
Response time = **days**

E-Mail
CRM SW
Input management

"human in the loop"

ERP-SW
Digital forms
Content Management

3. SW-based process
Response time = **hours**

First chatbots
Customer Analytics
Risk prediction

Rule-based SW
Planning Analytics
Digital workflows

4. AI-based process
Response time = **seconds**

Smart assistants
Customer models
Risk models

"human on the loop"

Financial models
Automated decisions
Automated workflows

Figure 3.1: Example of a typical transformation process towards AI-supported automation: the bank use case "loan for homeowners".

Cycles: AI development and operation

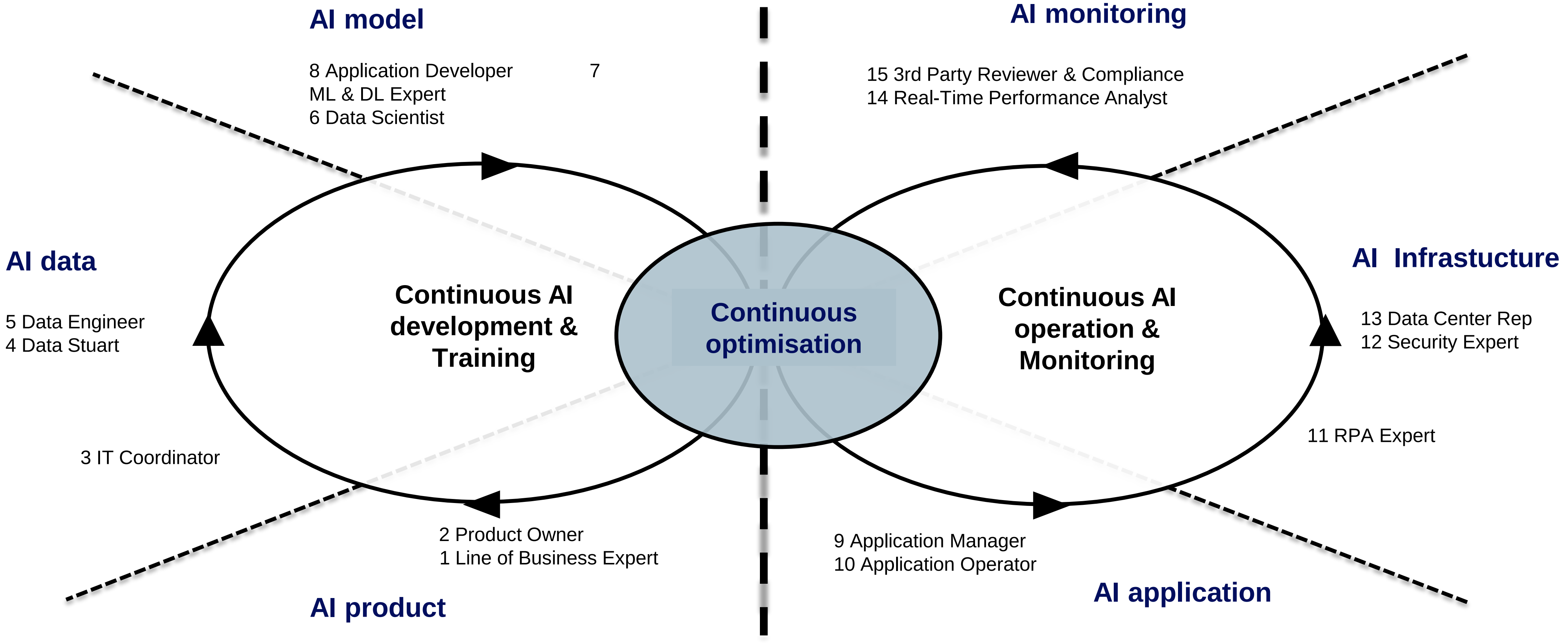
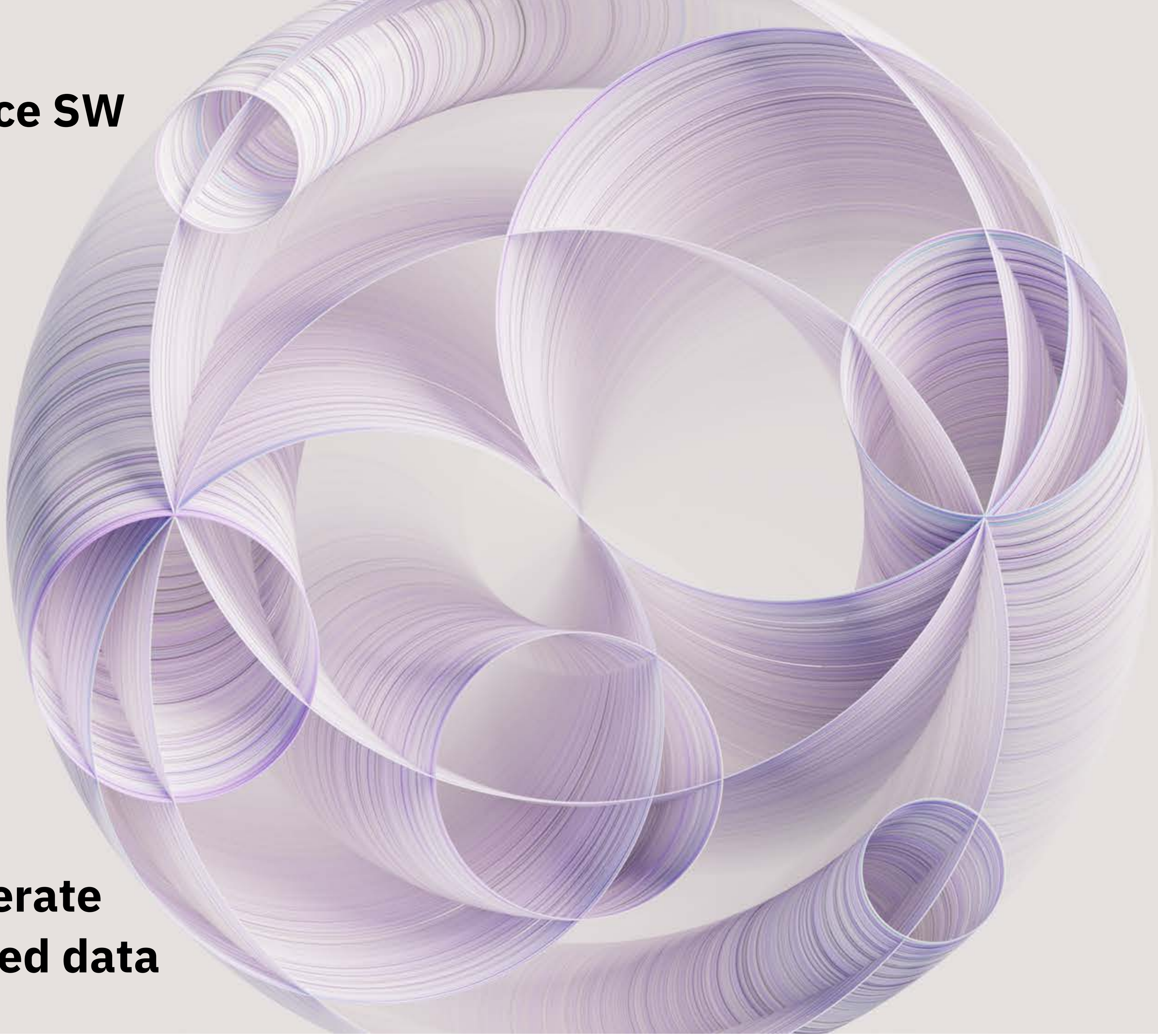


Figure 3.2: Life cycle of an AI application and relevant workflows.

IBM watsonx & open source SW

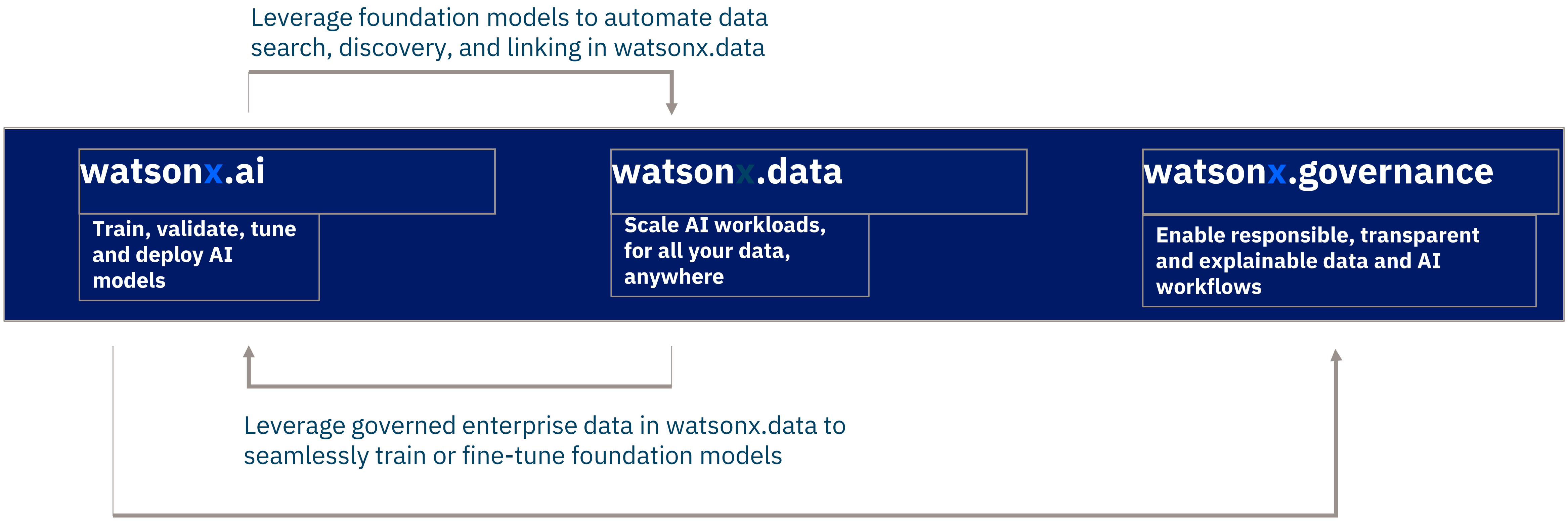
Elements of the platform



Scale, test & run
AI & Foundation Models
customize them and accelerate
the impact of AI with trusted data

Put AI to work with watsonx

Scaling and accelerating the impact of AI – including FM`s - with trusted data.



Leverage foundation models to automate data search, discovery, and linking in watsonx.data

watsonx.ai

Train, validate, tune and deploy AI models

watsonx.data

Scale AI workloads, for all your data, anywhere

watsonx.governance

Enable responsible, transparent and explainable data and AI workflows

Leverage governed enterprise data in watsonx.data to seamlessly train or fine-tune foundation models

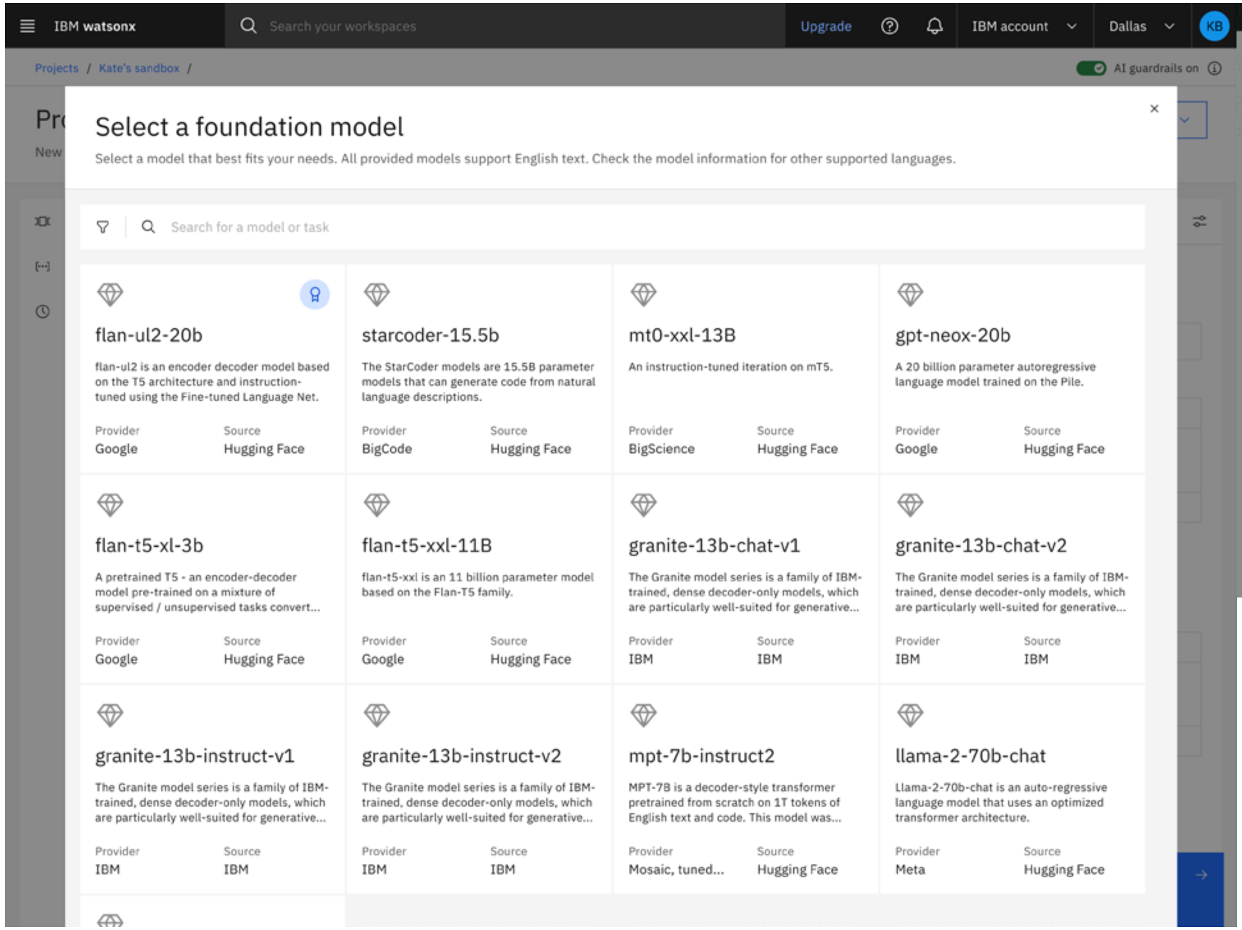
watsonx + Hugging Face

Enable fine-tuned models to be managed through market leading governance and lifecycle management capabilities

Watsonx.ai ↷

building, training, validating, tuning and deploying AI models

Multi-model & Multi-Cloud



IBM watsonx | Search your workspaces | Upgrade | IBM account | Dallas | KB

Projects / Kate's sandbox / | AI guardrails on

Select a foundation model

Select a model that best fits your needs. All provided models support English text. Check the model information for other supported languages.

Search for a model or task

flan-ul2-20b
flan-ul2 is an encoder decoder model based on the T5 architecture and instruction-tuned using the Fine-tuned Language Net.
Provider: Google | Source: Hugging Face

starcoder-15.5b
The StarCoder models are 15.5B parameter models that can generate code from natural language descriptions.
Provider: BigCode | Source: Hugging Face

mt0-xxl-13B
An instruction-tuned iteration on mT5.
Provider: BigScience | Source: Hugging Face

gpt-neox-20b
A 20 billion parameter autoregressive language model trained on the Pile.
Provider: Google | Source: Hugging Face

flan-t5-xl-3b
A pretrained T5 - an encoder-decoder model pre-trained on a mixture of supervised / unsupervised tasks convert...
Provider: Google | Source: Hugging Face

flan-t5-xxl-11B
flan-t5-xxl is an 11 billion parameter model based on the Flan-T5 family.
Provider: Google | Source: Hugging Face

granite-13b-chat-v1
The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative...
Provider: IBM | Source: IBM

granite-13b-chat-v2
The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative...
Provider: IBM | Source: IBM

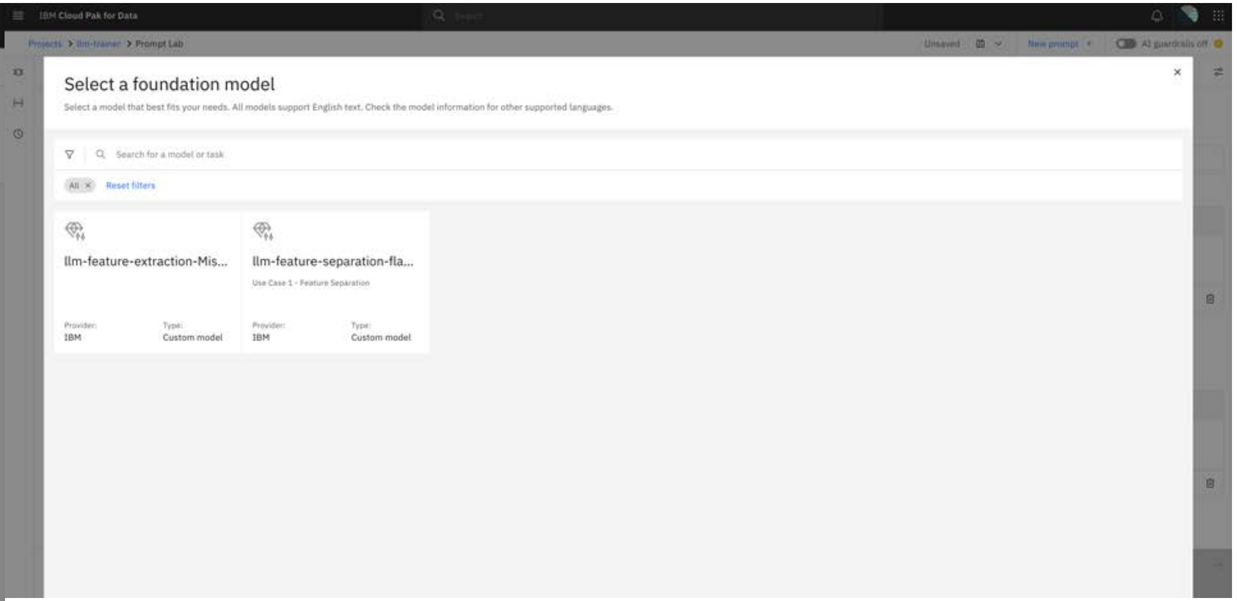
granite-13b-instruct-v1
The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative...
Provider: IBM | Source: IBM

granite-13b-instruct-v2
The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative...
Provider: IBM | Source: IBM

mpt-7b-instruct2
MPT-7B is a decoder-style transformer pretrained from scratch on 1T tokens of English text and code. This model was...
Provider: Mosaic, tuned... | Source: Hugging Face

llama-2-70b-chat
Llama-2-70b-chat is an auto-regressive language model that uses an optimized transformer architecture.
Provider: Meta | Source: Hugging Face

Own customized models & MLOps



IBM Cloud Pak for Data

Projects > llm-trainer > Prompt Lab | Unsaved | New prompt | AI guardrails off

Select a foundation model

Select a model that best fits your needs. All models support English text. Check the model information for other supported languages.

Search for a model or task

All ✕ | Reset filters

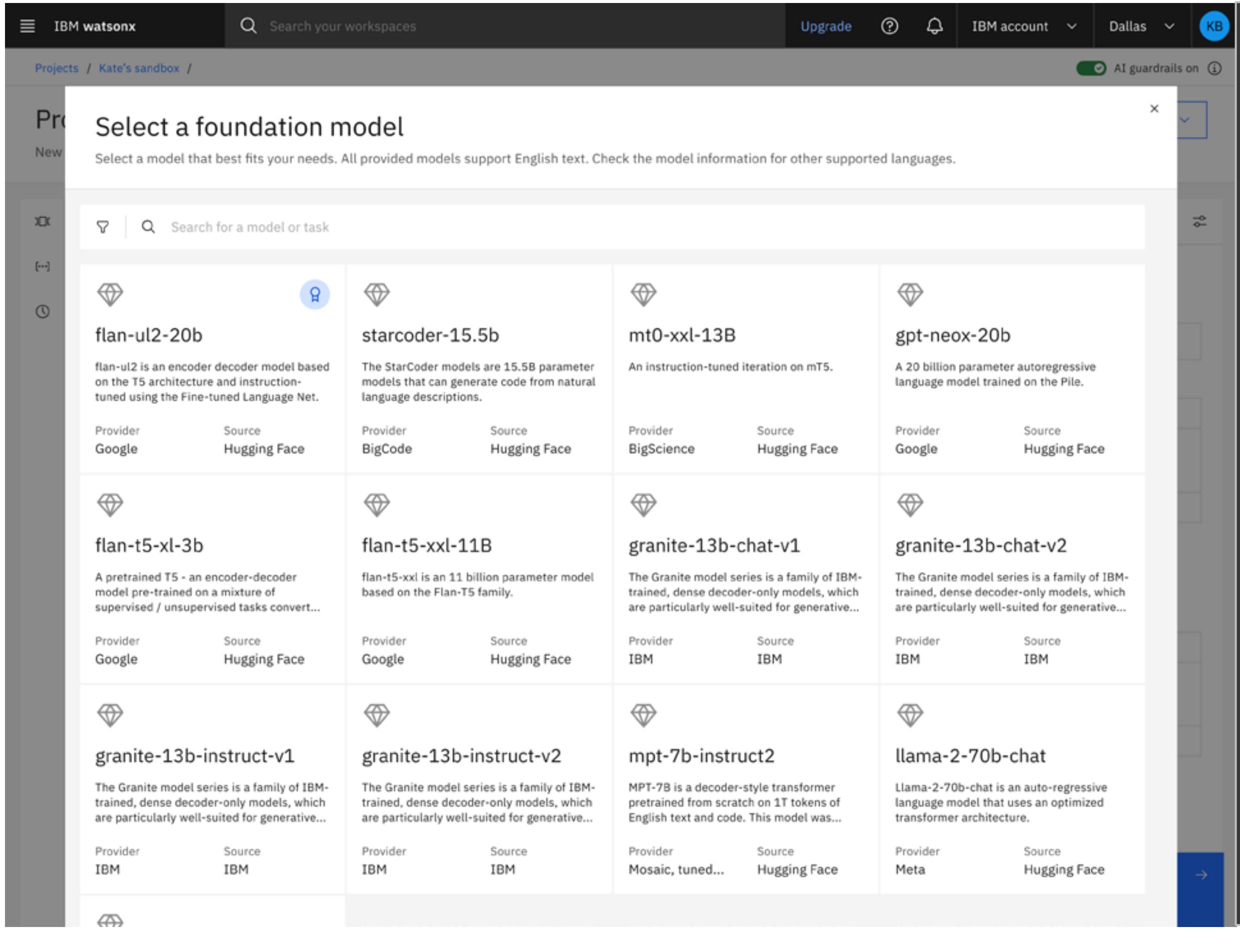
llm-feature-extraction-Mis...
Provider: IBM | Type: Custom model

llm-feature-separation-fla...
Use Case 1 - Feature Separation
Provider: IBM | Type: Custom model

Watsonx.ai

building, training, validating, tuning and deploying AI models

Multi-model & Multi-Cloud



Select a foundation model

Select a model that best fits your needs. All provided models support English text. Check the model information for other supported languages.

flan-ul2-20b
flan-ul2 is an encoder decoder model based on the T5 architecture and instruction-tuned using the Fine-tuned Language Net.
Provider: Google — Source: Hugging Face

starcoder-15.5b
The StarCoder models are 15.5B parameter models that can generate code from natural language descriptions.
Provider: BigCode — Source: Hugging Face

mt0-xxl-13B
An instruction-tuned iteration on mT5.
Provider: BigScience — Source: Hugging Face

gpt-neox-20b
A 20 billion parameter autoregressive language model trained on the Pile.
Provider: Google — Source: Hugging Face

flan-t5-xl-3b
A pretrained T5 - an encoder-decoder model pre-trained on a mixture of supervised / unsupervised tasks convert...
Provider: Google — Source: Hugging Face

flan-t5-xxl-11B
flan-t5-xxl is an 11 billion parameter model based on the Flan-T5 family.
Provider: Google — Source: Hugging Face

granite-13b-chat-v1
The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative...
Provider: IBM — Source: IBM

granite-13b-chat-v2
The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative...
Provider: IBM — Source: IBM

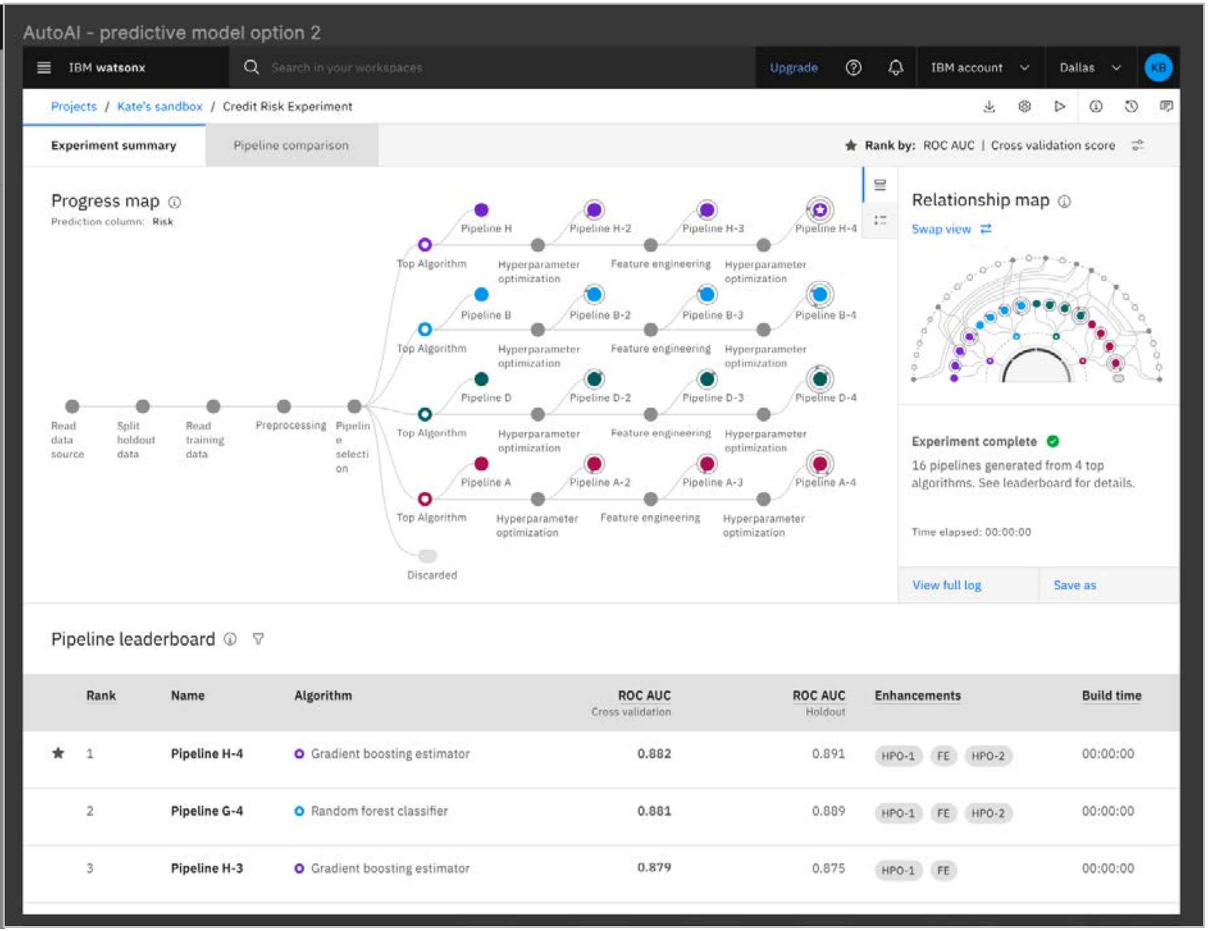
granite-13b-instruct-v1
The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative...
Provider: IBM — Source: IBM

granite-13b-instruct-v2
The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative...
Provider: IBM — Source: IBM

mpt-7b-instruct2
MPT-7B is a decoder-style transformer pretrained from scratch on 1T tokens of English text and code. This model was...
Provider: Mosaic, tuned... — Source: Hugging Face

llama-2-70b-chat
Llama-2-70b-chat is an auto-regressive language model that uses an optimized transformer architecture.
Provider: Meta — Source: Hugging Face

Data Science & MLOps



Experiment complete

16 pipelines generated from 4 top algorithms. See leaderboard for details.

Time elapsed: 00:00:00

Pipeline leaderboard

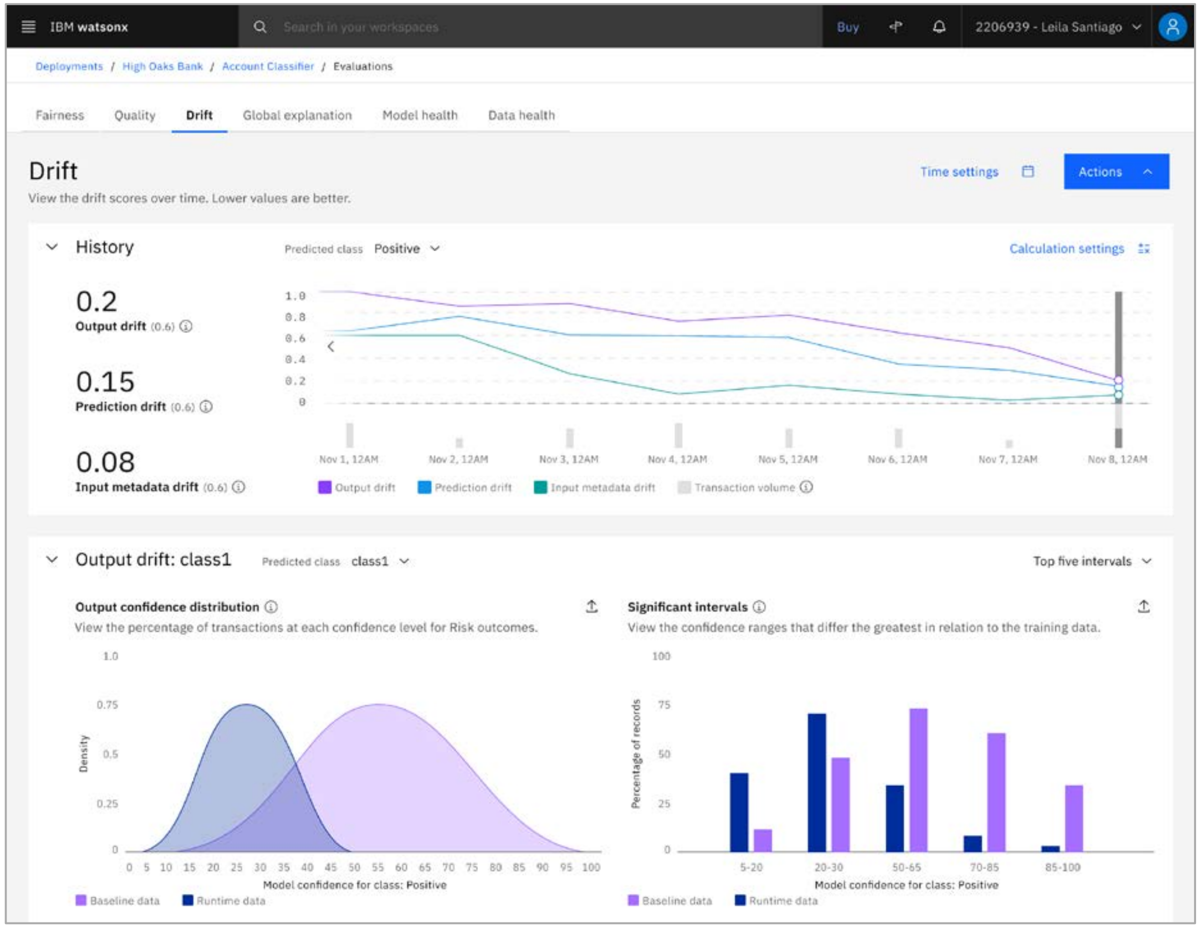
	Rank	Name	Algorithm	ROC AUC Cross validation	ROC AUC Holdout	Enhancements	Build time
★	1	Pipeline H-4	Gradient boosting estimator	0.882	0.891	HPO-1 FE HPO-2	00:00:00
	2	Pipeline G-4	Random forest classifier	0.881	0.889	HPO-1 FE HPO-2	00:00:00
	3	Pipeline H-3	Gradient boosting estimator	0.879	0.875	HPO-1 FE	00:00:00

Watsonx.governance

controlling the performance of your models & AI governance via dashboards

DIN

Evaluation and monitoring



IBM watsonx

Search in your workspaces

Buy

2206939 - Leila Santiago

Deployments / High Oaks Bank / Account Classifier / Evaluations

Fairness | Quality | **Drift** | Global explanation | Model health | Data health

Drift

View the drift scores over time. Lower values are better.

Time settings

Actions

History

Predicted class: Positive

Calculation settings

0.2
Output drift (0.6)

0.15
Prediction drift (0.6)

0.08
Input metadata drift (0.6)

Output drift: class1

Predicted class: class1

Top five intervals

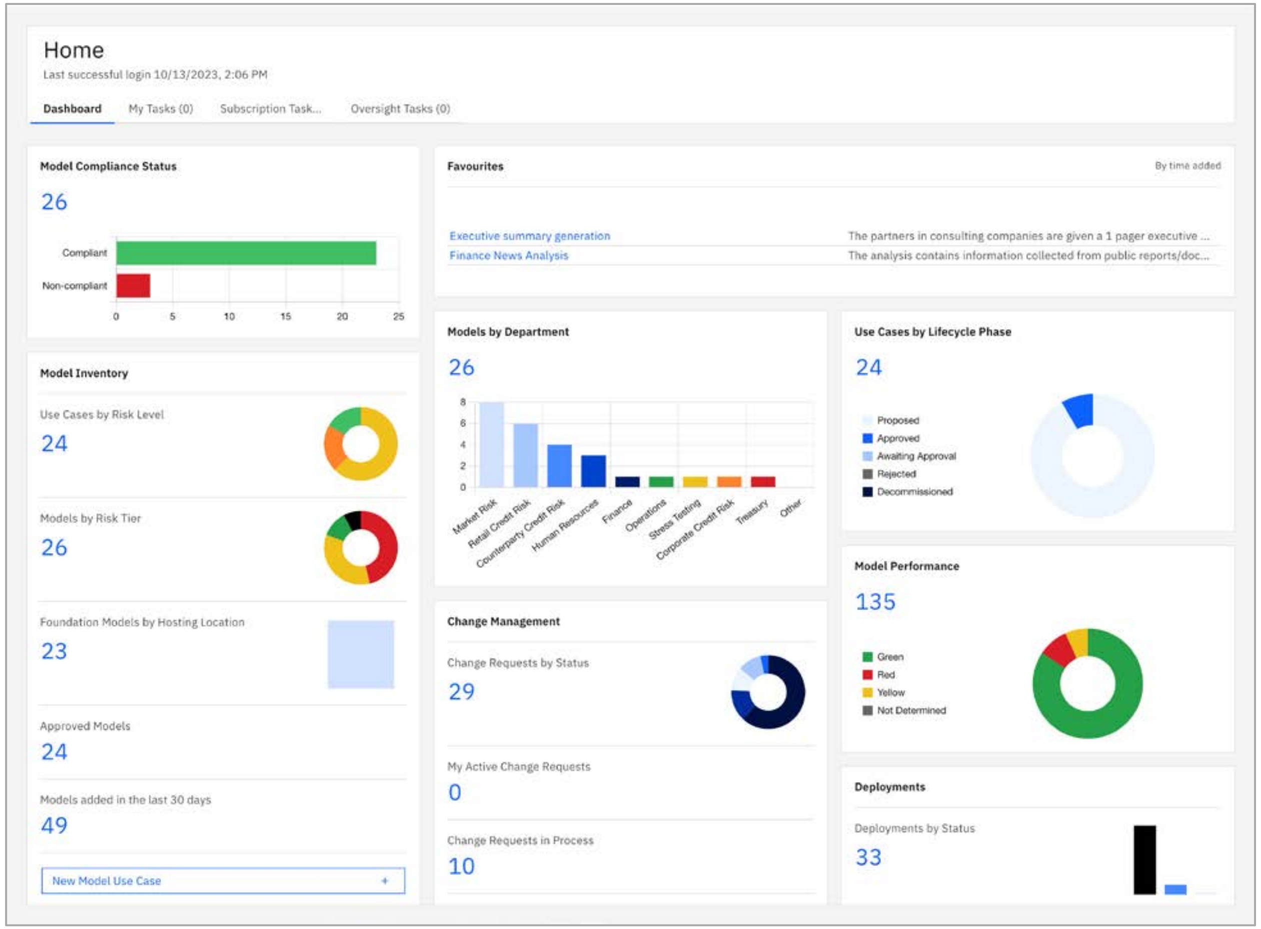
Output confidence distribution

View the percentage of transactions at each confidence level for Risk outcomes.

Significant intervals

View the confidence ranges that differ the greatest in relation to the training data.

Manage risk and best practices



Home

Last successful login 10/13/2023, 2:06 PM

Dashboard | My Tasks (0) | Subscription Task... | Oversight Tasks (0)

Model Compliance Status: 26

Favourites — By time added

Executive summary generation	The partners in consulting companies are given a 1 pager executive ...
Finance News Analysis	The analysis contains information collected from public reports/doc...

Model Inventory

Use Cases by Risk Level: 24

Models by Risk Tier: 26

Foundation Models by Hosting Location: 23

Approved Models: 24

Models added in the last 30 days: 49

New Model Use Case +

Models by Department: 26

Use Cases by Lifecycle Phase: 24

Change Management

Change Requests by Status: 29

My Active Change Requests: 0

Change Requests in Process: 10

Model Performance: 135

Deployments

Deployments by Status: 33

IBM

Important aspects of an AI ecosystem from any Organisation

Every societal partner should participate:

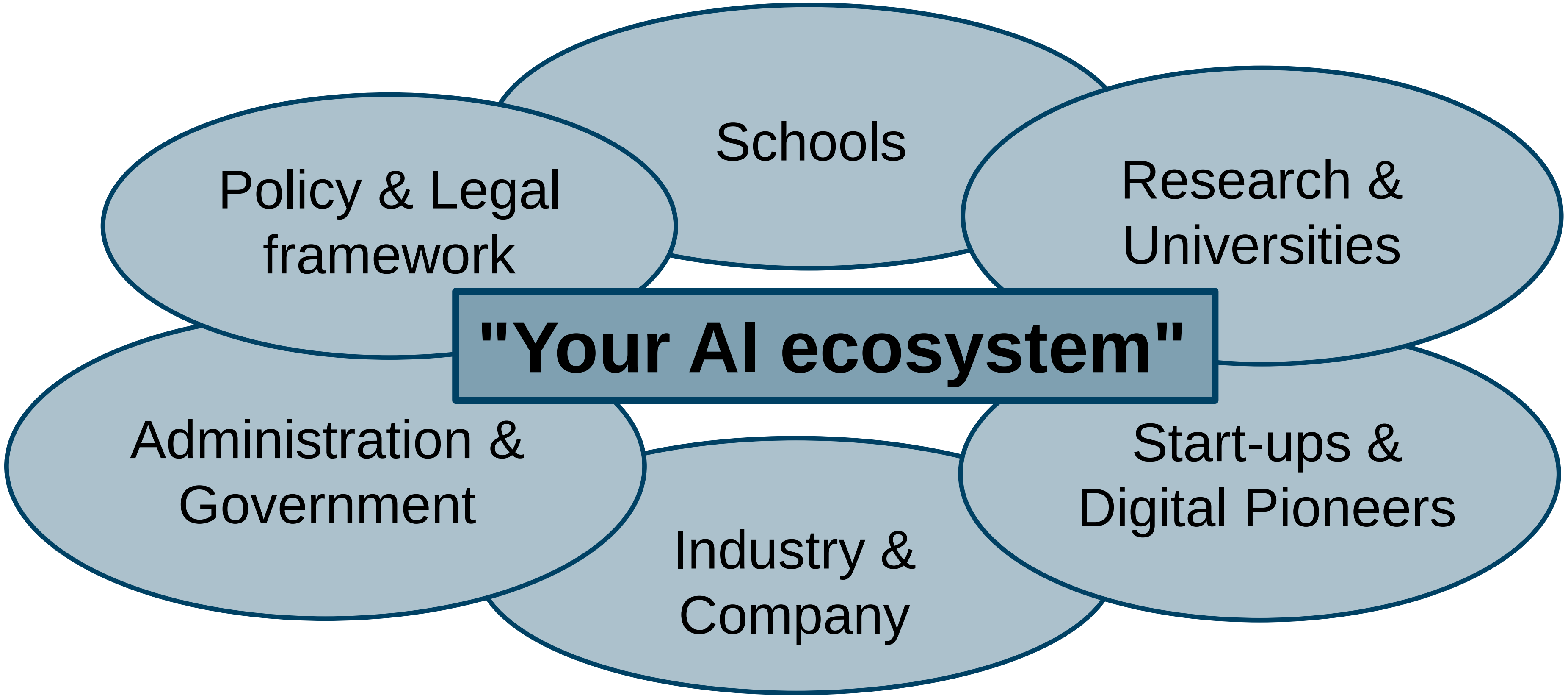


Figure 4.3: An AI ecosystem should involve every societal partner,that matters.

Hochrisikoanwendung & Testing

Wenn Sie eine High-Risk KI-Anwendung haben, wie testen Sie diese?

	Herausforderungen/Probleme	Chancen/Vorteile
Vor Betrieb		
Im Betrieb		
Nach Betrieb ´		

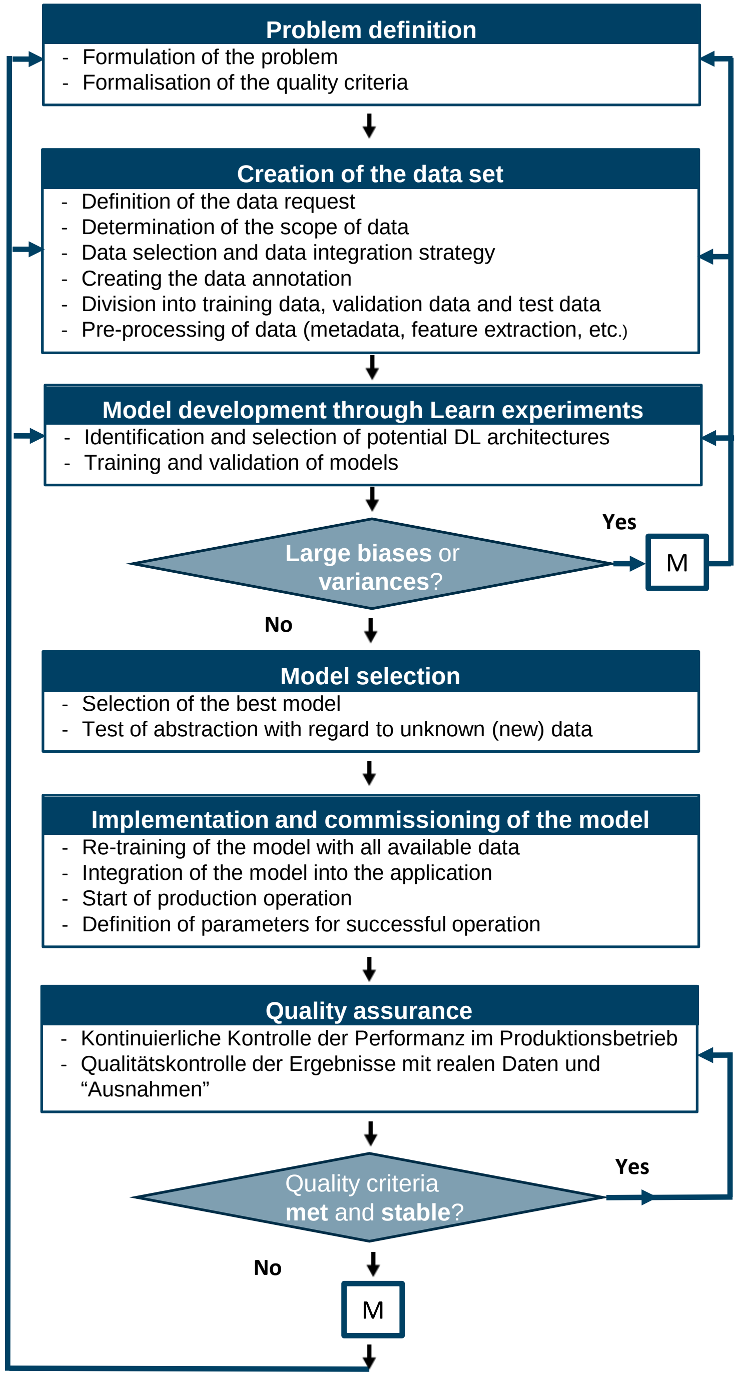
4. Example of Application of the EU-AI Act using AI norms to create advantage

Image recognition in medicine
Help with diagnosis

Figure 18.2: AI-assisted analysis of medical image data.

Image recognition in medicine
Help with diagnosis

Figure 18.3: Flow chart for the development of a deep learning systemof a deep learning systemaccording to **DIN SPEC 13266**.



Problem definition

- Formulation of the problem
- Formalisation of the quality criteria

Creation of the data set

- Definition of the data request
- Determination of the scope of data
- Data selection and data integration strategy
- Creating the data annotation
- Division into training data, validation data and test data
- Pre-processing of data (metadata, feature extraction, etc.)

Model development through Learn experiments

- Identification and selection of potential DL architectures
- Training and validation of models

Large biases or **variances**?

Yes → M

No

Model selection

- Selection of the best model
- Test of abstraction with regard to unknown (new) data

Implementation and commissioning of the model

- Re-training of the model with all available data
- Integration of the model into the application
- Start of production operation
- Definition of parameters for successful operation

Quality assurance

- Kontinuierliche Kontrolle der Performanz im Produktionsbetrieb
- Qualitätskontrolle der Ergebnisse mit realen Daten und "Ausnahmen"

Quality criteria **met** and **stable**?

Yes

No → M

M **Management decision**

Q **Regular quality assurance loop**

4 – KI Normen als Wettbewerbsvorteil

Was muss passieren, damit die Nutzung von KI-Normung Wettbewerbsvorteile bieten kann?

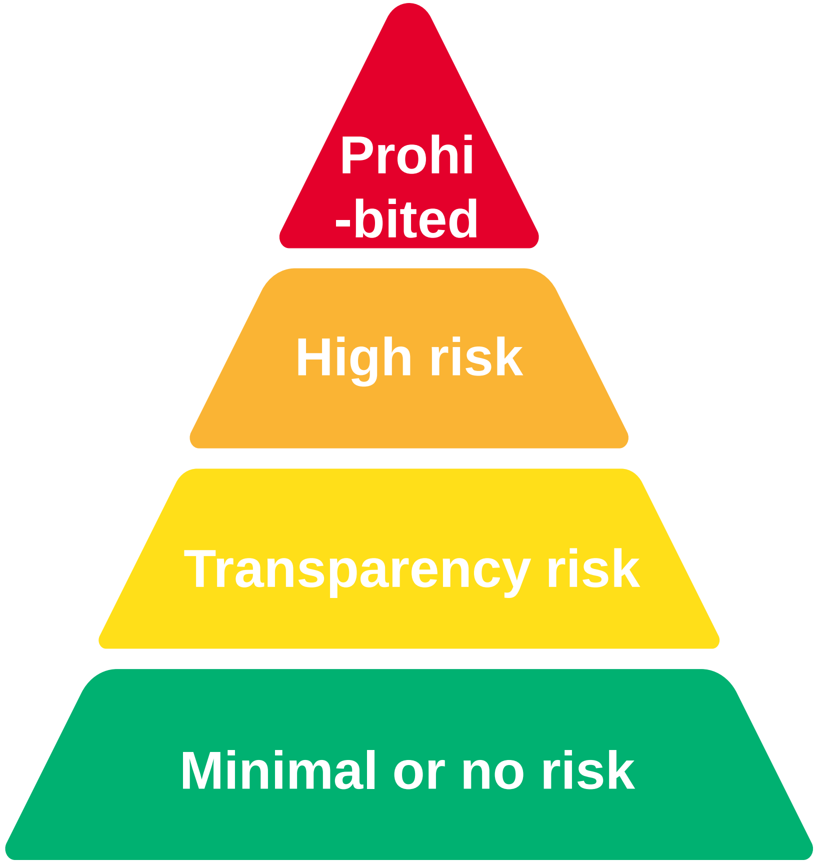
Welche offenen/ungelösten Fragen gibt es zum EU AI Act?

5. Implications of risk class according to the EU AI Act

Prohibited

Examples

- Systems that deploy subliminal techniques to manipulate users.
- AI systems used for social scoring by public authorities.
- AI systems used for real-time biometric identification in public spaces (with certain exceptions for law enforcement).



Due Diligence of the adopter/user

Adopters are not allowed to use or adopt these systems at all.

Any engagement with such systems is illegal under the AI Act.

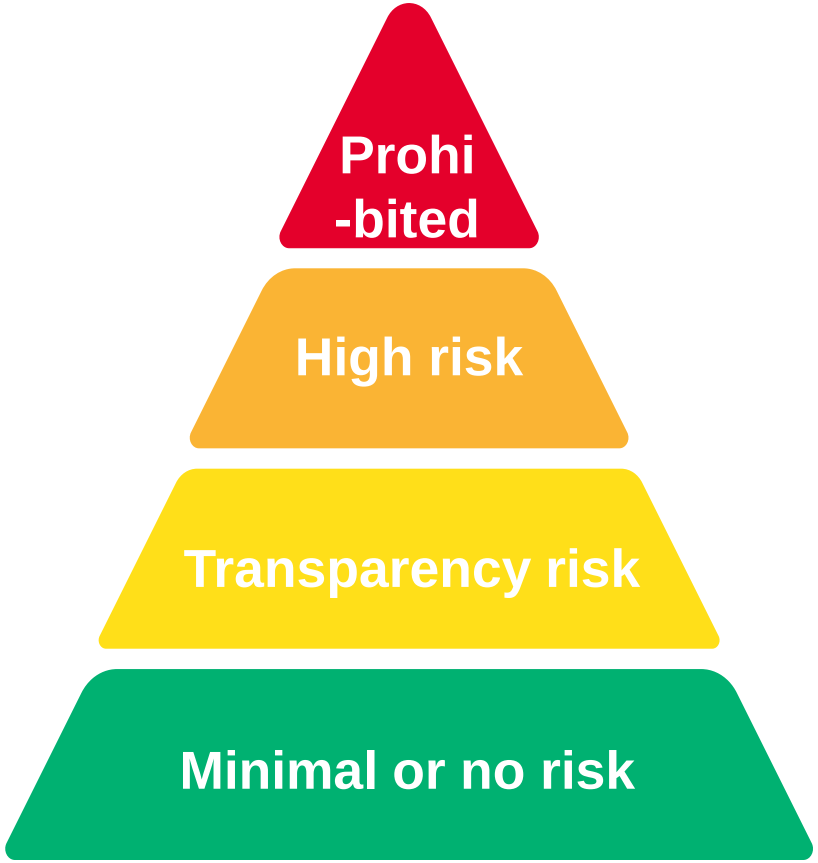
Organisations must ensure that they do not inadvertently deploy prohibited systems by conducting thorough compliance checks.

High Risk

DIN

AI used in

- critical infrastructure
- Education
- Employment
- essential public and private services
- law enforcement
- biometric identification



Due Diligence of the adopter/user

Risk Management: Establish systems to assess and mitigate risks throughout the AI lifecycle.

Data Governance: Ensure high-quality, bias-free datasets and maintain data integrity.

Transparency & Documentation: Keep detailed records on the system’s design, development, and compliance.

Human Oversight: Implement measures for human intervention and monitoring.

Post-Market Monitoring: Continuously monitor performance to ensure ongoing compliance.

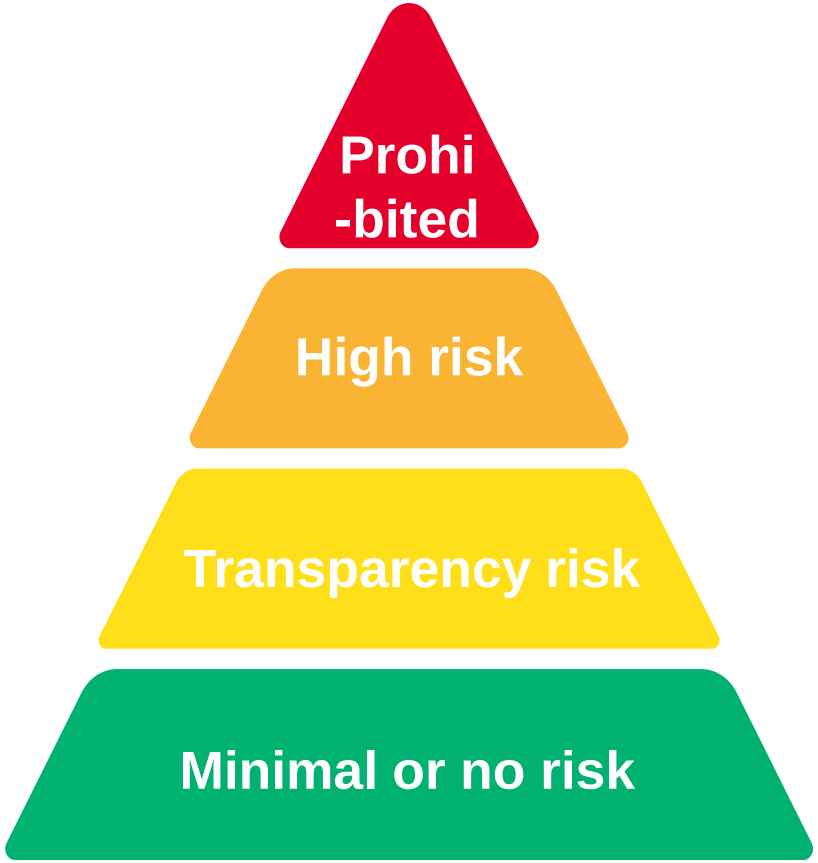
Notified Body Involvement:

- Conformity Assessment: Engage a notified body for independent audit and certification.
- Certification & Surveillance: Obtain certification and undergo ongoing compliance checks.

Transparency Risk

Examples

- AI chatbots or systems that interact with humans.
- Systems that generate deepfakes (unless they are clearly labeled).



Due Diligence of the adopter/user

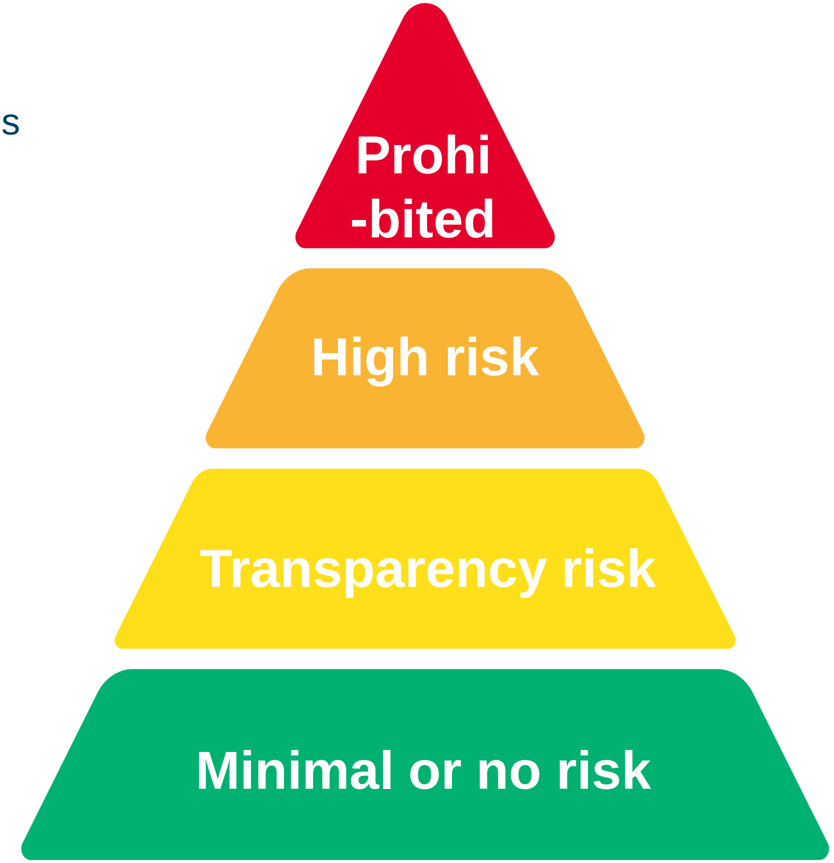
Transparency: Adopters must inform users when they are interacting with an AI system. For instance, a chatbot should explicitly state that it is AI-driven.

User Awareness: Ensure users understand that they are engaging with AI and provide necessary information for users to make informed decisions about the interaction.

Minimal or No Risk

Examples

- AI systems used in games or email filters.



Due Diligence of the adopter/user

Minimal Regulatory Obligation: No specific regulatory requirements or due diligence beyond general legal obligations (such as data protection under GDPR).

How to prepare for the EU AI Act

- Pledge to fully **embrace** the use of AI in the business.
- **Delegate sufficient authority** in regards to the management of the AI system.
- **Classify the risk level** of the AI system given its anticipated use.
- **Adhere** to all internal and external regulatory policies.
- Consider **specific risks** of the business aligned with AI application.
- Maintain **risk management** and **quality assurance** in an expanded manner.
- Focus on your competitive advantage provided by your AI systems
- Use conformity assessments and declarations to accelerate your EU GTM and sales.

KI managen und verstehen

Vertrauen durch Transparenz

Wie schafft man Vertrauen in KI?

Lesen Sie dazu jetzt den neuen Praxis-Wegweiser **Künstliche Intelligenz managen und verstehen.**

› Hier mehr

Thomas Schmid, Wolfgang Hildesheim, Taras Holoyad

DIN

Künstliche Intelligenz managen und verstehen

Der Praxis-Wegweiser für Entscheidungsträger, Entwickler und Regulierer

Beuth

https://www.beuth.de/de/themenseiten/ki