WHEN TRUST MATTERS

AI og praktisk anvendelse

Borekonferansen 2025

Christian Markussen (Global Practice Lead – Technology Qualification & Digital Twins)

29 October 2025

Evolution of AI

- **Machine learning** – (1950's) algorithms that are trained on large datasets to find patterns and **make predictions**
- **Perception AI** – (2012) ability to interpret, understand and **classify data**, images, and video
- **Generative AI** – (2020) ability to understand the meaning of the information and **translate (generate)** it into different languages, images, code, etc.
- **Agentic AI** – (2024) ability to autonomously **pursue goals** and complete complex multi-step tasks using **planning and reasoning**
- **Physical AI** – (near future) ability to understand and **reason in the physical world**, e.g. understand the law of physics, object permanence, cause and effect, predicting outcomes, estimating how much force is needed to grip an object without damage, etc.

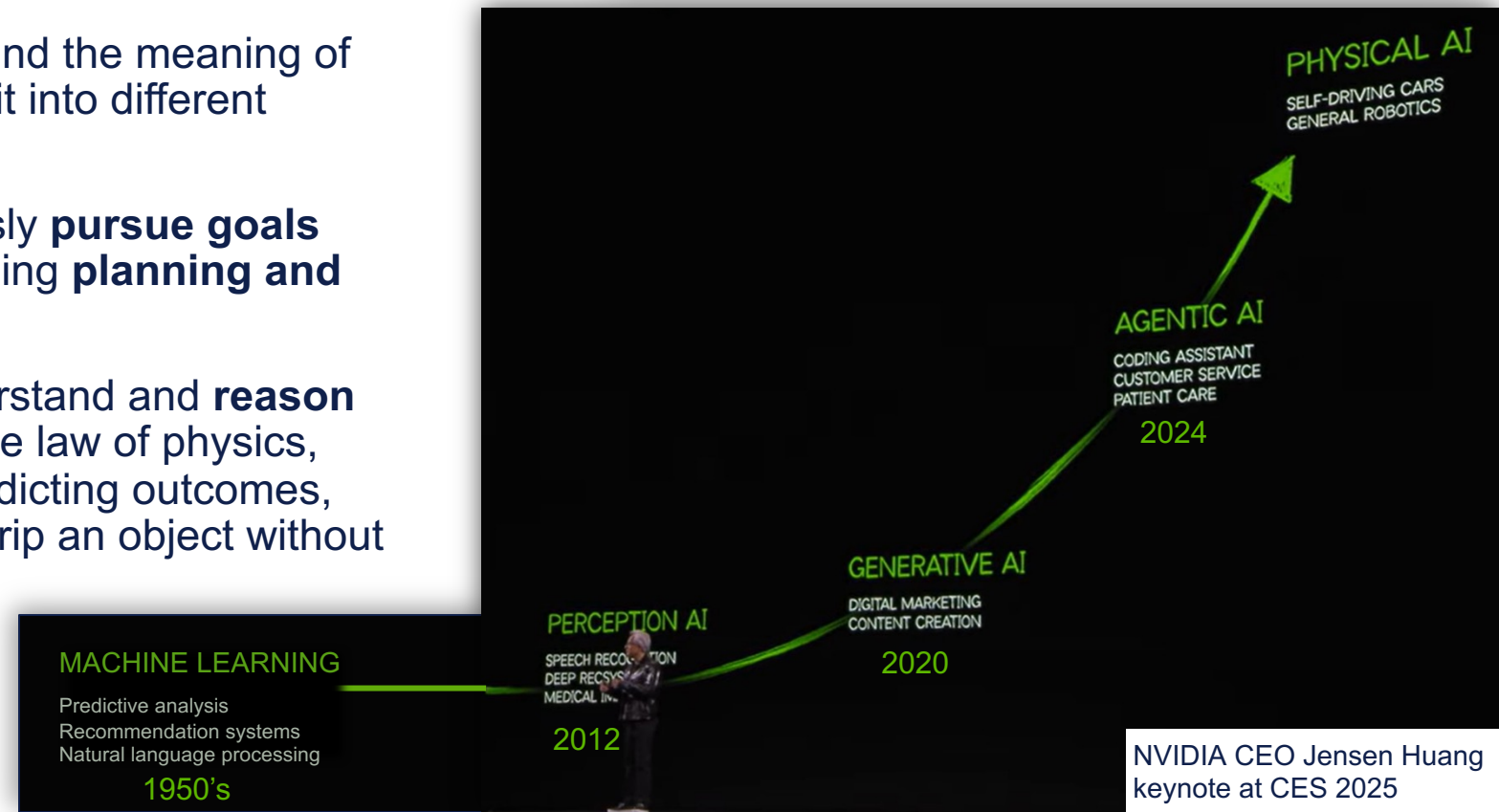
IBM Agentic AI: 4 reasons why it's the next big thing in AI research

Subscribe To Newsletters Forbes

LEADERSHIP > CMO NETWORK

How Agentic AI Will Turn Your Life And Workplace Upside-down

TIME How the Rise of New Digital Workers Will Lead to an Unlimited Age



NVIDIA CEO Jensen Huang keynote at CES 2025

We are in an AI hype bobble

E24 | Norges største næringslivsavis

Hvorfor uteblir gevinsten av AI-satsingen?

Fortsatt er det få bedrifter som har økt avkastningen fordi de tar i bruk kunstig intelligens. Vi har forsket på hva som skal til for å lykkes.

AI Promises Productivity. It's Delivering 'Workslop'

October 12, 2025 at 10:00 PM GMT+2

By Catherine Thorbecke

Catherine Thorbecke is a Bloomberg Opinion columnist covering Asia tech. Previously she was a tech reporter at CNN and ABC News.

Futurism

HEAD IN THE CLOUD

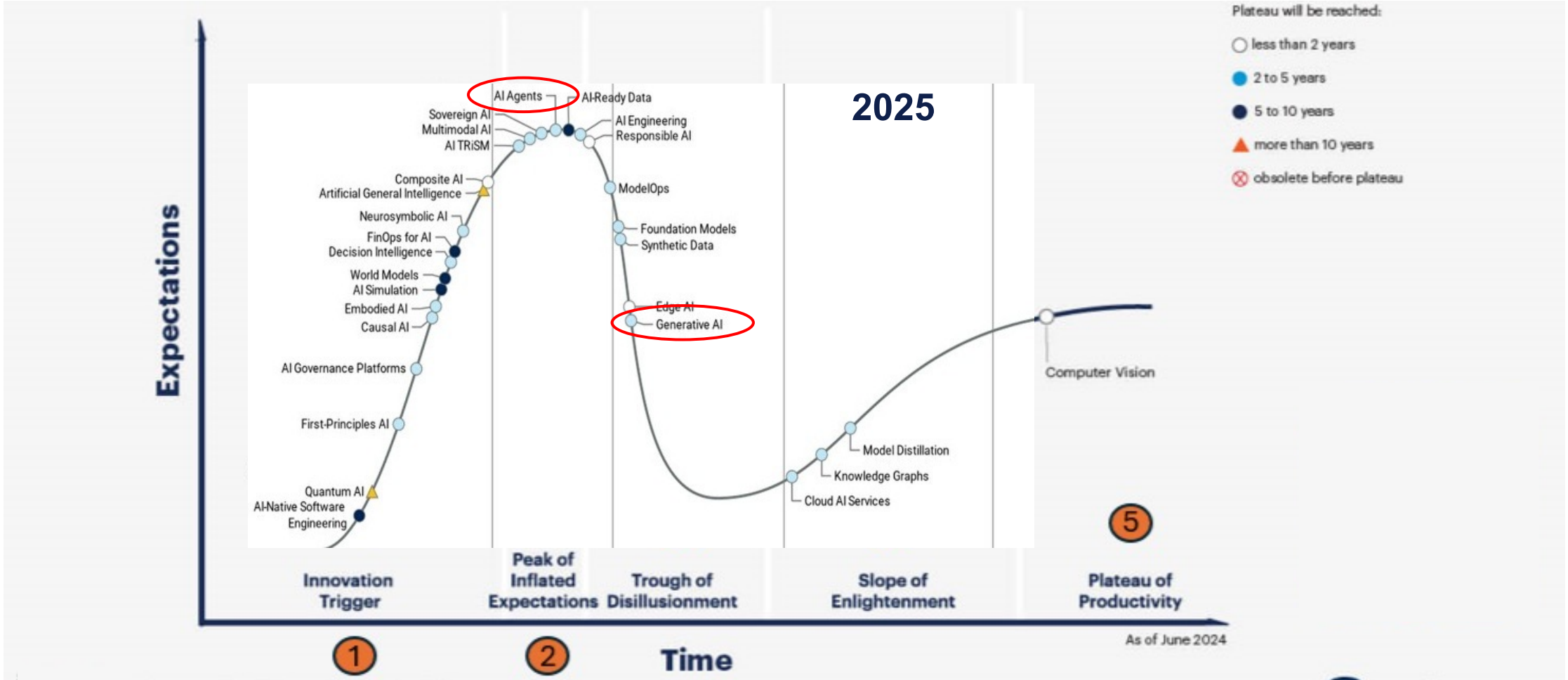
AWS Outage That Took Down Internet Came After Amazon Fired Tons of Workers in Favor of AI

ITPro.

AI-generated code is now the cause of one-in-five breaches – but developers and security leaders alike are convinced the technology will come good eventually

DNV

Hype cycle for Artificial Intelligence

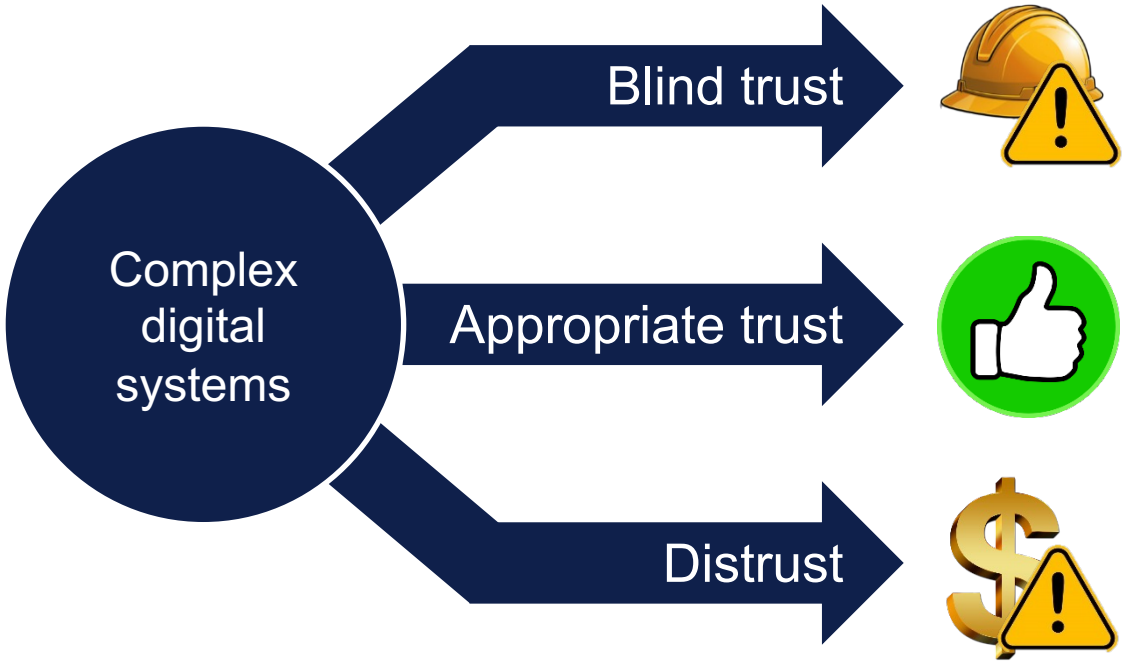


Curtesy of Gartner

DNV

Trust in complex systems

- Trust: the belief that someone or something is reliable, truthful, and will act in one's best interest



Misuse, over-confidence ➔ Unsafe or costly actions

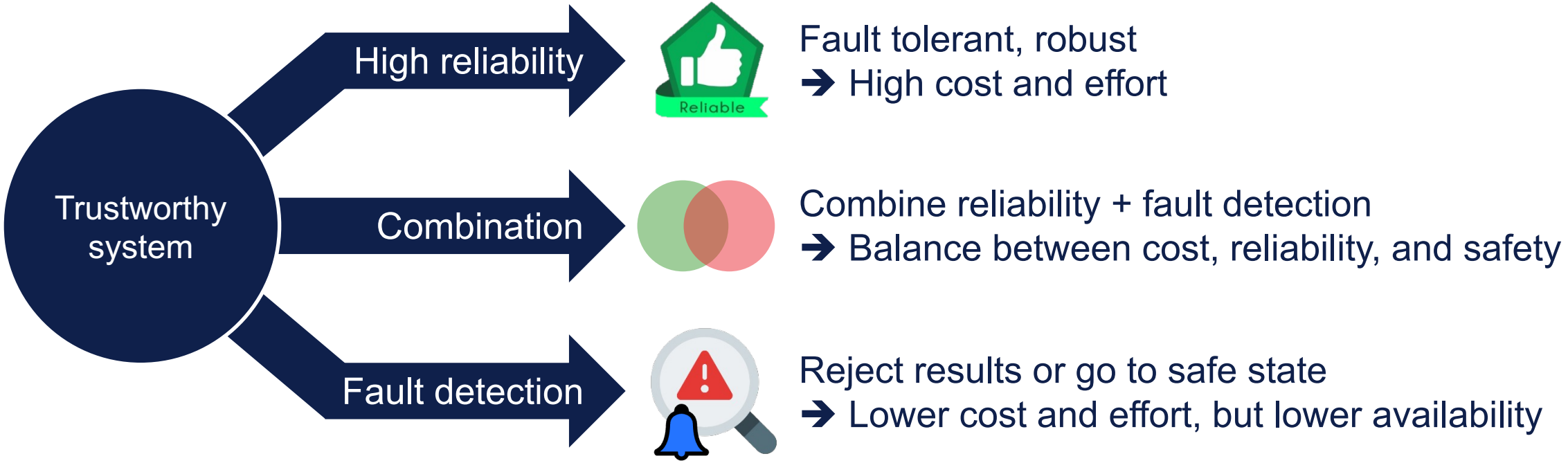
Trustworthy system ➔ Valuable and safe

Disuse, resistance ➔ Undermines business case

Complex system: a system whose behaviour can not be understood from its individual components

Strategy for building a trustworthy system

- Generative AI-based systems are generally unreliable
  - ~80% correctness / reliability is easily achievable
  - >99% correctness / reliability is difficult and costly



Human interaction

"We have trained operators so they will be able to manage the situation"

"It is not safety critical since we have a human-in-the-loop"

However,

the more capable and complex the system is, the less people are able to …

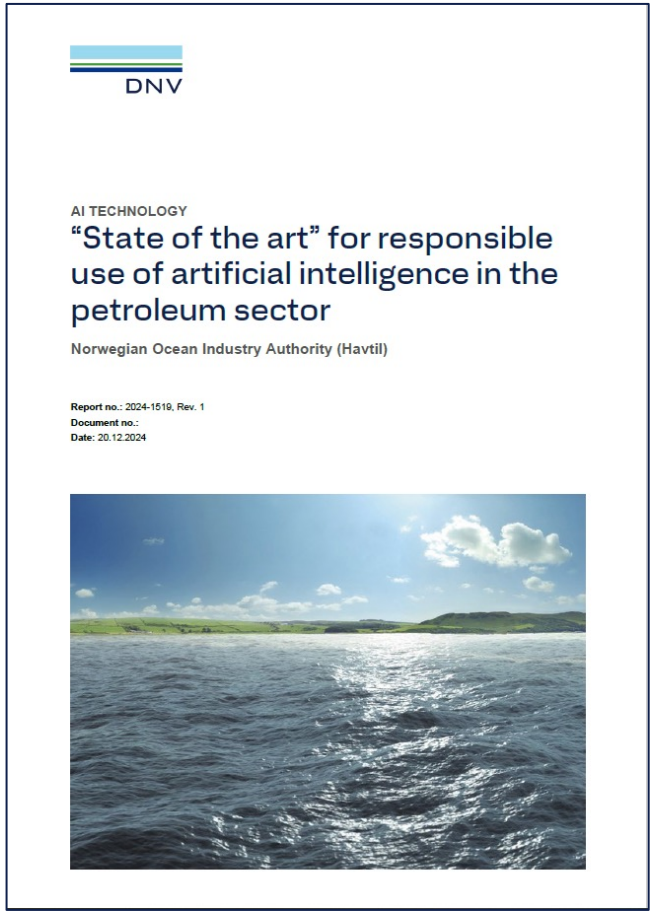
- understand it
- understand its limits
- compensate for its shortcomings

How does the user, or down-stream system, know if the information from the AI-based system can be trusted?

Source: Endsley et.al.

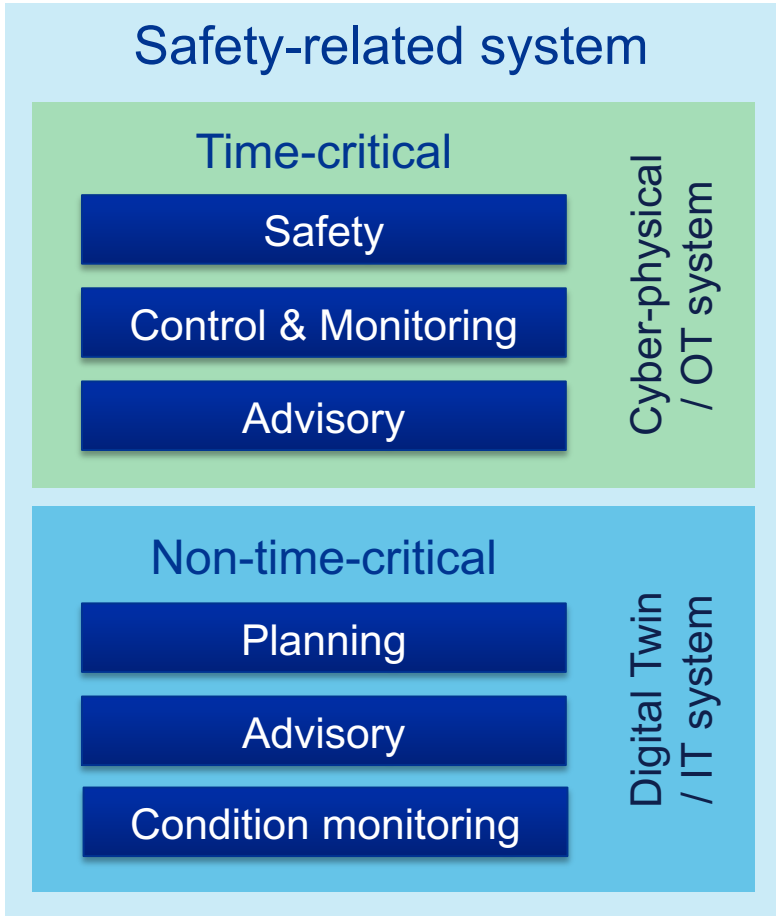
Havtil "Letter on artificial intelligence follow-up"

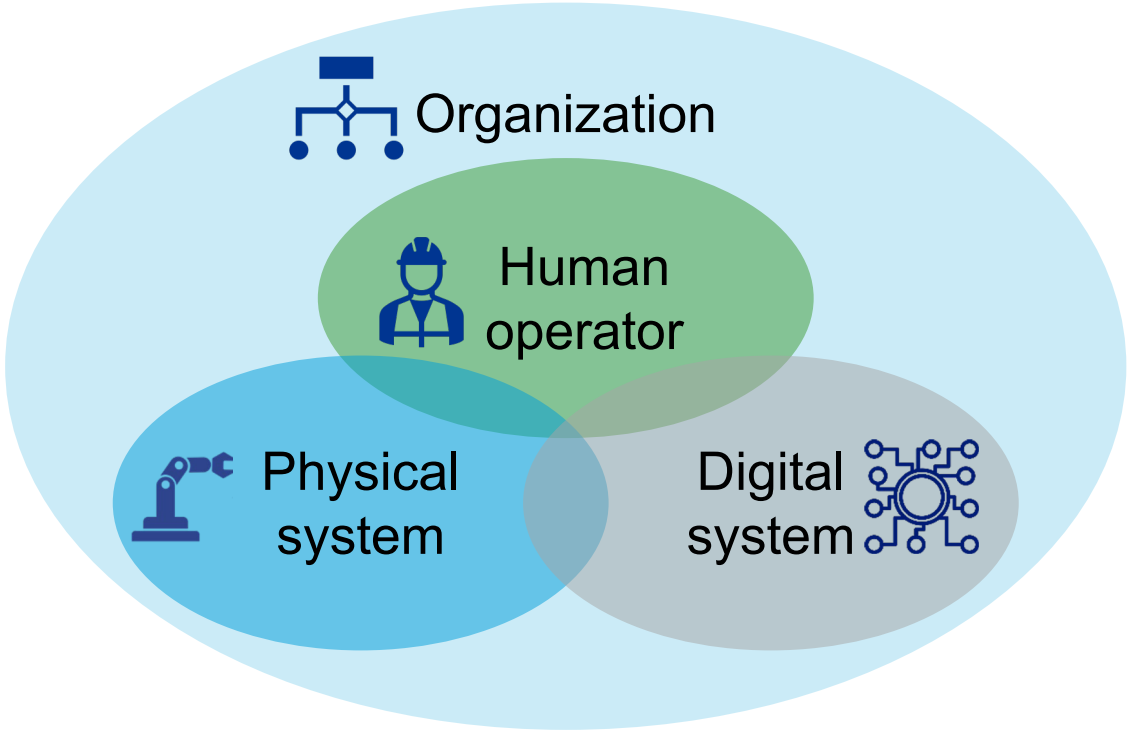
- Havtil guidance on 15 regulation § relevant for AI-based (and complex digital) systems:
  - Supervision and human oversight
  - Risk management & uncertainty
  - Transparency, explainability & uncertainty
  - Lifecycle approach
  - Documentation & auditability
  - Employee involvement & competence
  - Cybersecurity & robustness



Source: Havtil

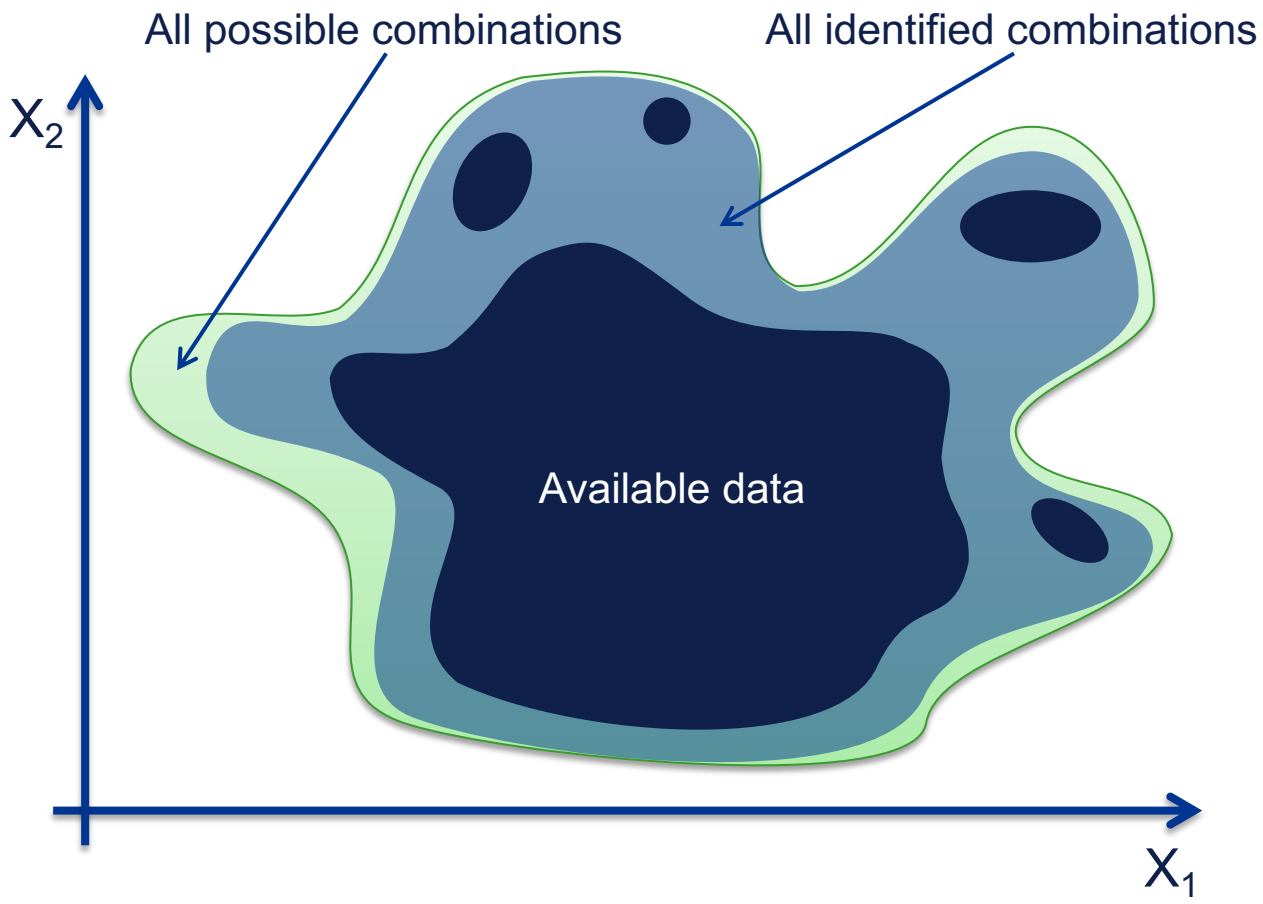
Responsible use of AI





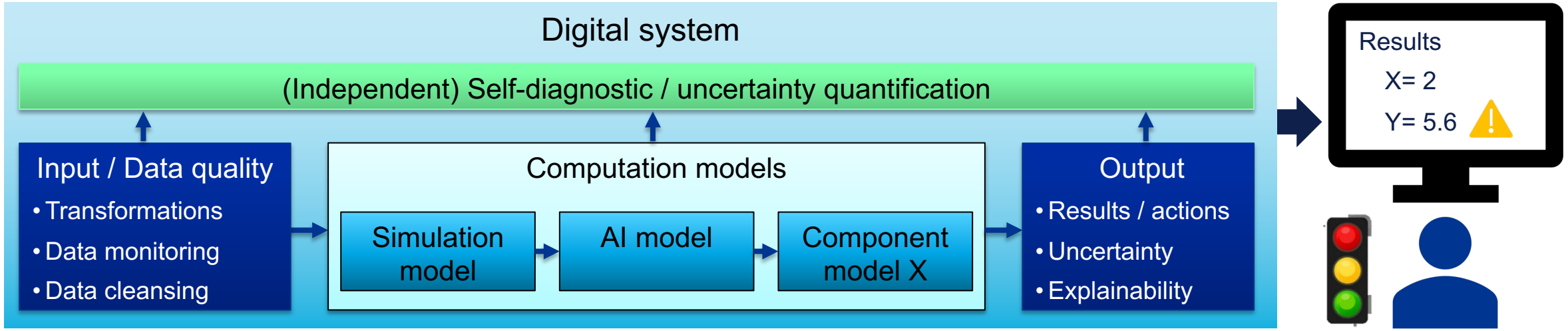
Source: DNV report for Havtil: "State of the art" for responsible use of artificial intelligence in the petroleum sector

Data and training of AI

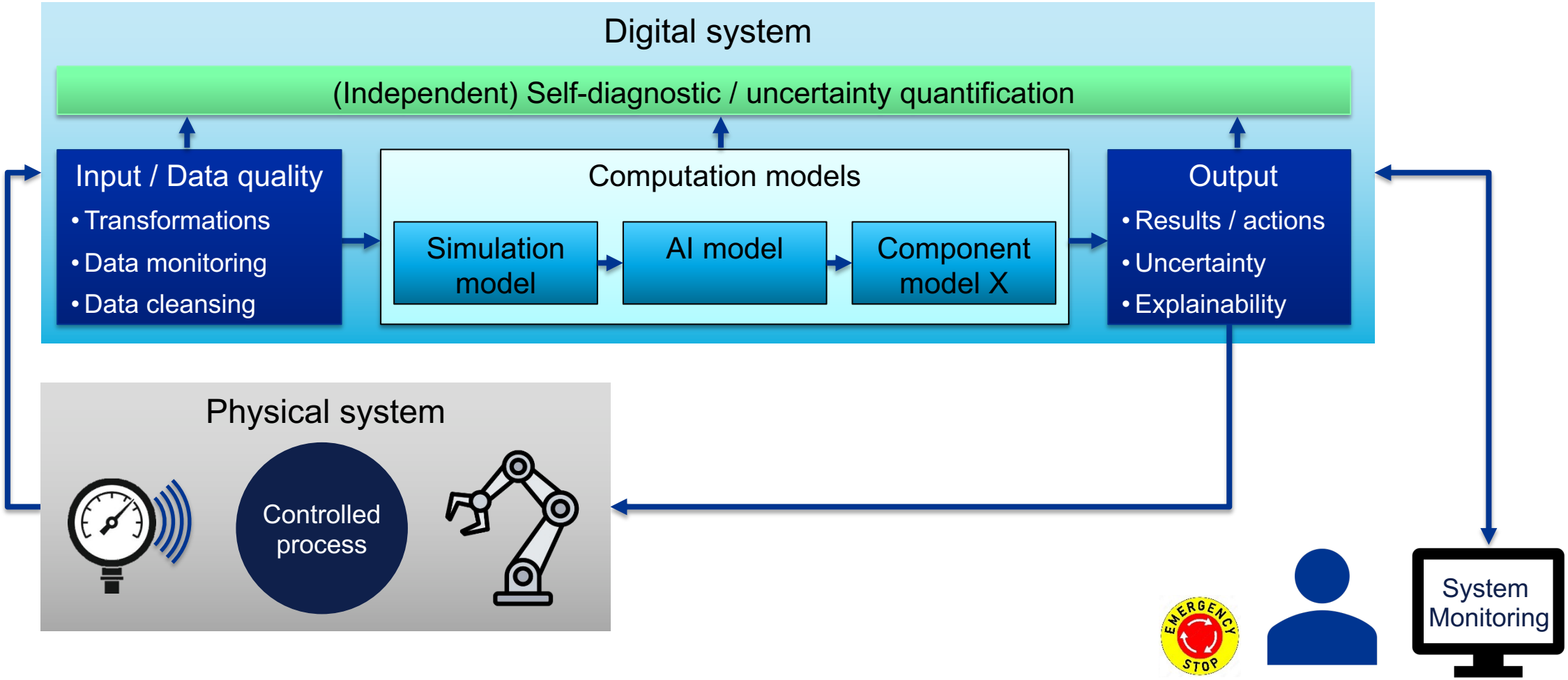


- Training data is a bug, not a feature
  - Data is a **proxy for knowledge** and expertise
  - We have a lot of data, but typically lack data for edge cases and anomalies
  - The training data must contain the patterns for the desired knowledge
- Augment the AI
  - Include **physics-based models**, e.g. synthetic data, hybrid models, constrained by physics, etc.
  - Include rules and **expert knowledge**
- Make the AI-system **uncertainty aware**
  - Uncertainty predictions
  - Independent self-assessment and health indicator

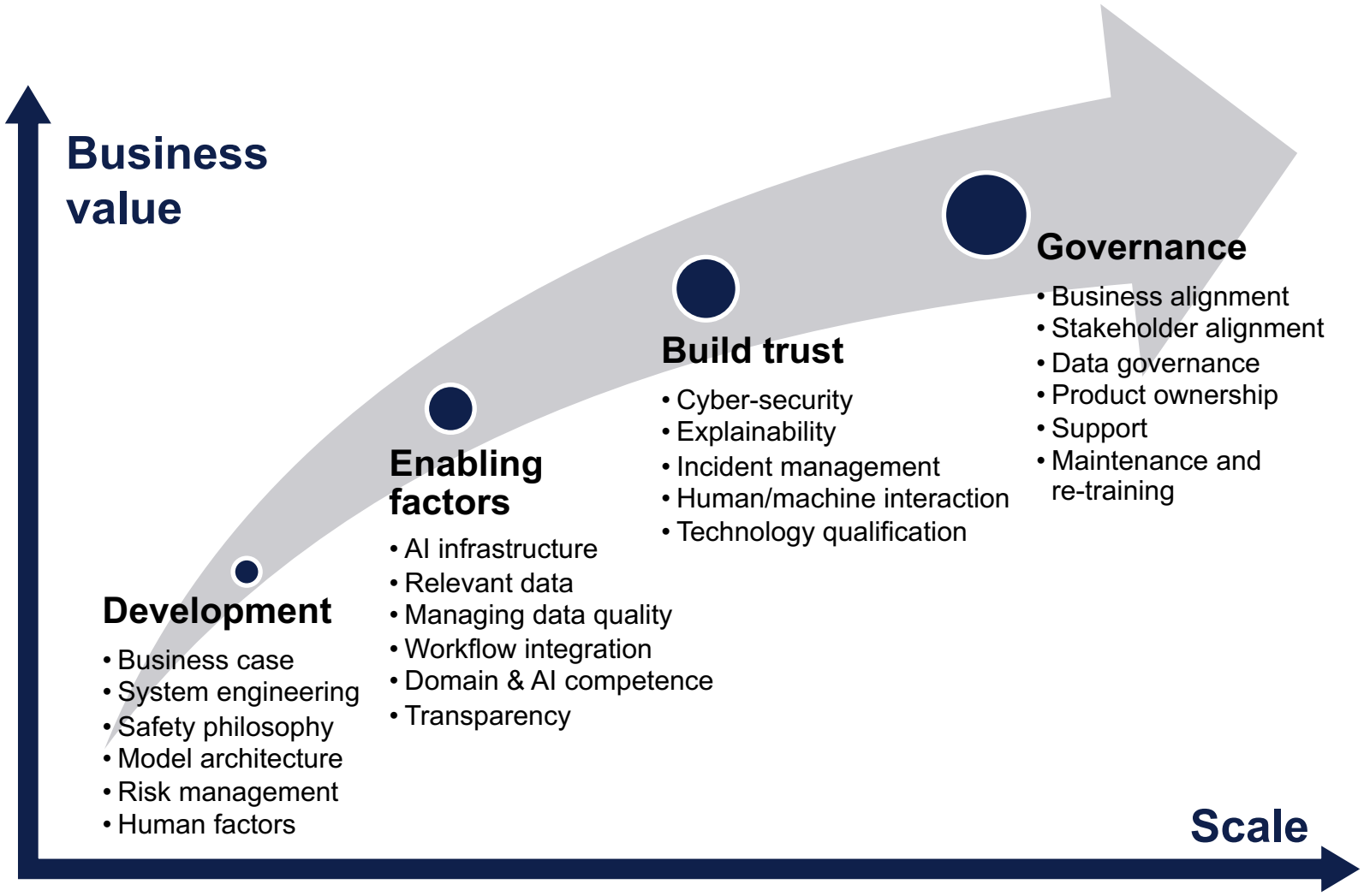
Trustworthy Digital Twin system (DNV-RP-A204)

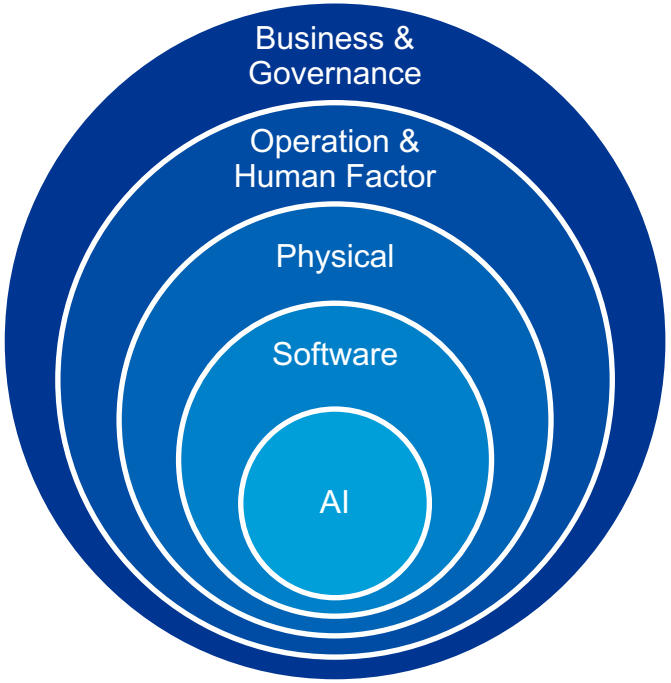


Trustworthy Cyber-Physical system (DNV-RP-A203)



Building business value





Summary

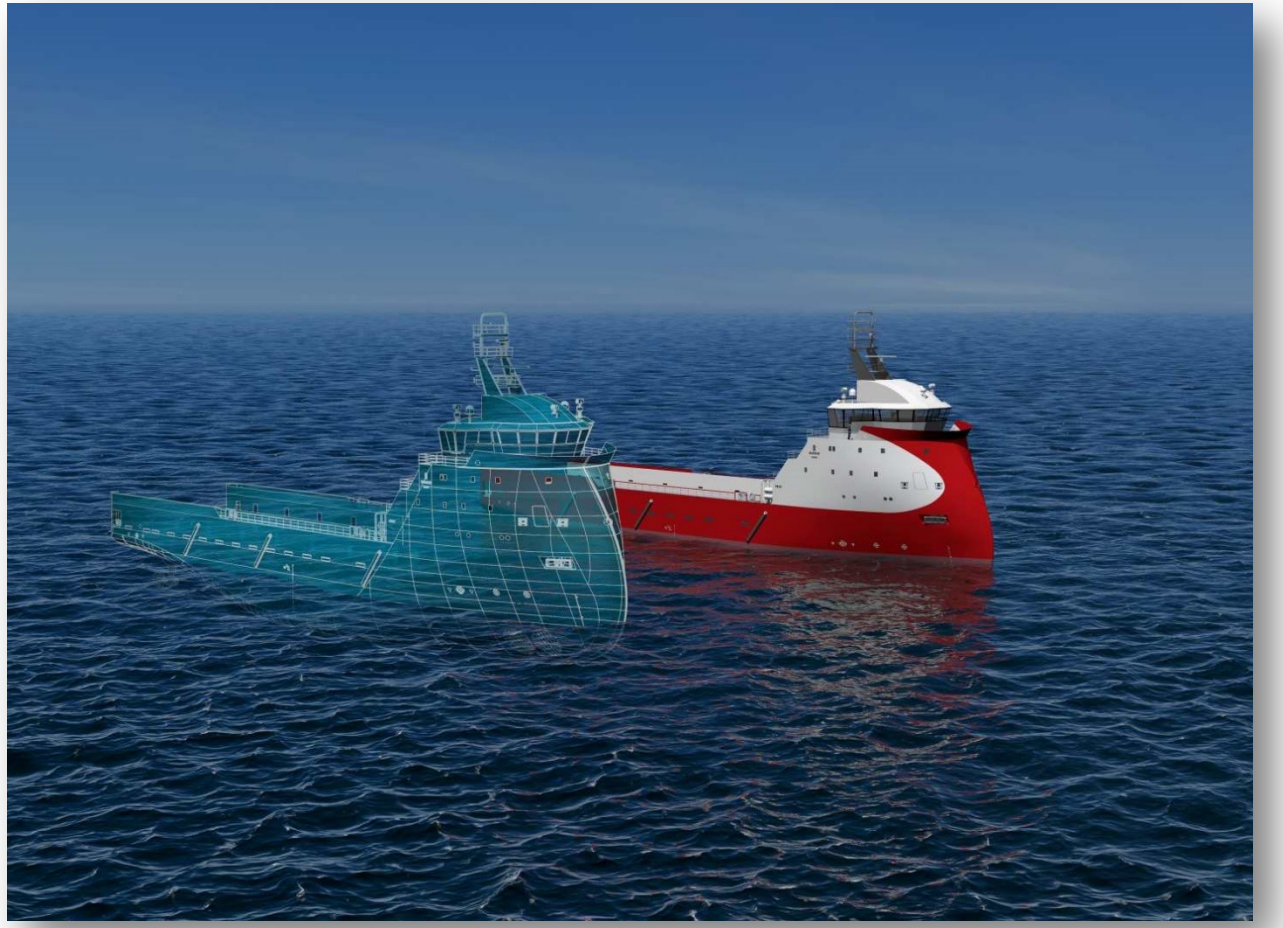
- **A lot of AI hype, but it is important to get ready**
  - Are you building capabilities, frameworks, and partners that can create trustworthy AI?
  - Can you quickly select the winning use cases and concepts?
  - Can you scale from PoC to a qualified system?
- **Trust is essential for both a solid business case and for safety**
  - Misinformation, leading to unsafe actions, is the biggest risk from AI
  - More capable and complex system ➔ Less ability to understand and supervise
  - Can the operator maintain situational awareness and act as an independent barrier?
- **Build trustworthiness into the AI solutions**
  - Bringing all knowledge into AI – not just data
  - AI cannot be practically qualified as a “black box” ➔ integrate the qualification process
  - Design system to detect and handle faults and communicate uncertainties to the user
- **Norges Rederiforbund** – Responsible use of AI in the O&G industry, Nov. 11th

DNV standards related to digital systems

Technology Qualification
RP-A203

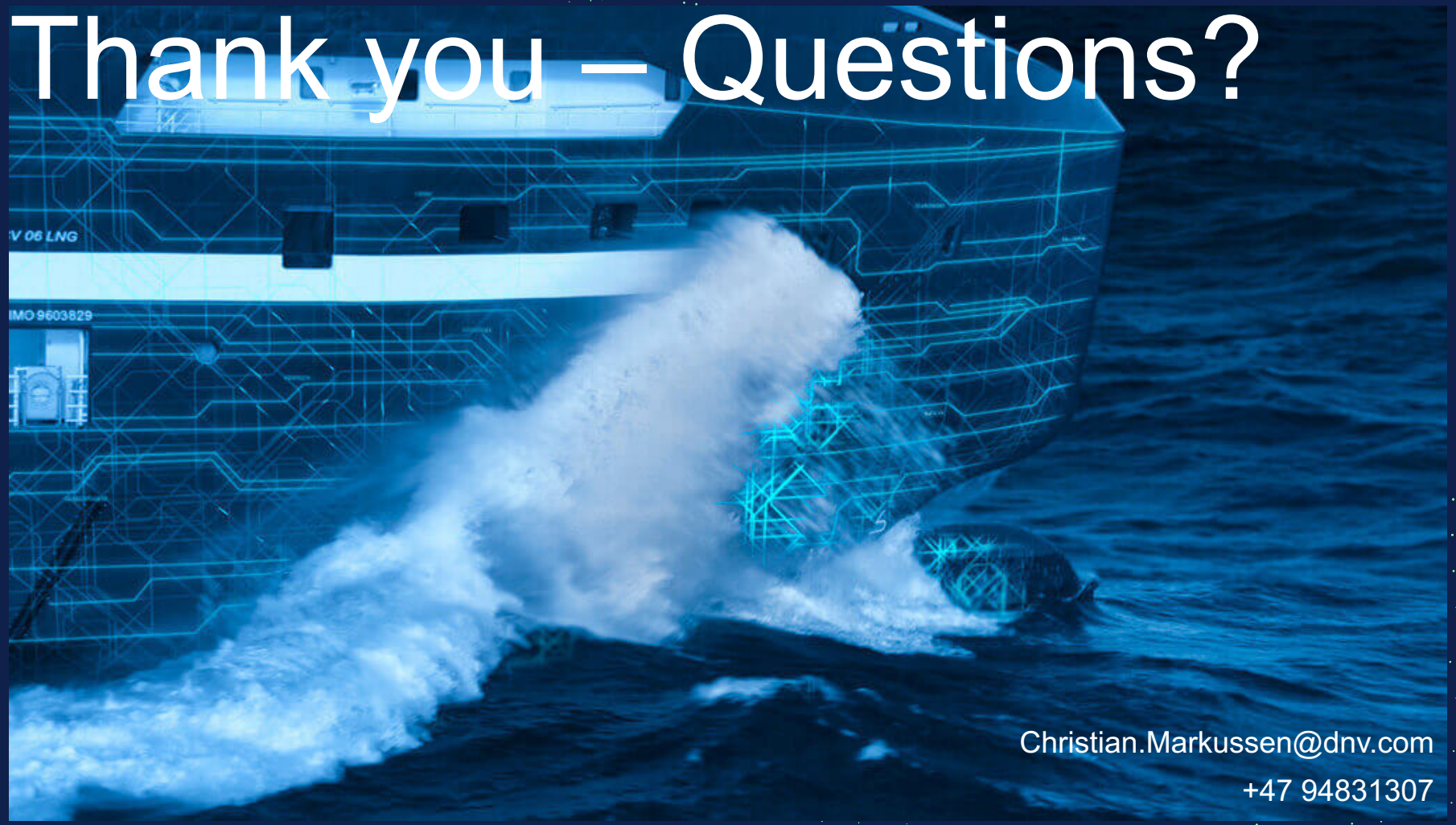
Assurance of Digital Twins
RP-A204

Data quality RP-0497	**Machine learning application** RP-0665	**Data-driven models** RP-0510	**Simulation models** RP-0513
Cyber security RP-G108 RP-0575 RP-0498 CP-0231	**Artificial intelligence** RP-0671	**Sensor systems** RP-0317	**Information modelling framework** RP-0670



Search for "DNV rules and standards explorer"

Thank you – Questions?



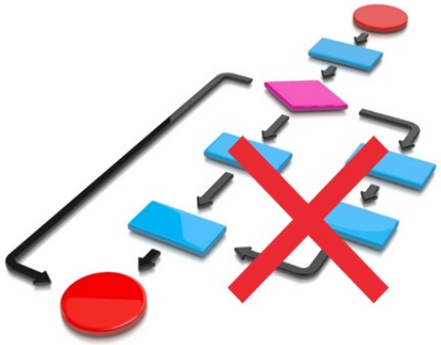
Christian.Markussen@dnv.com
+47 94831307

www.dnv.com

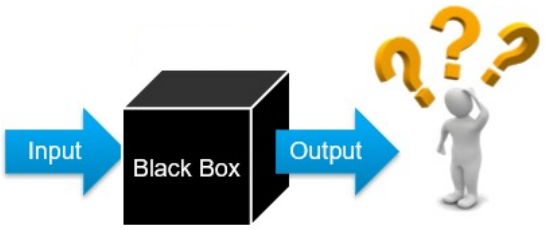
DNV

Why do some Digital Systems fail to deliver value?

Too technology oriented, not business case driven



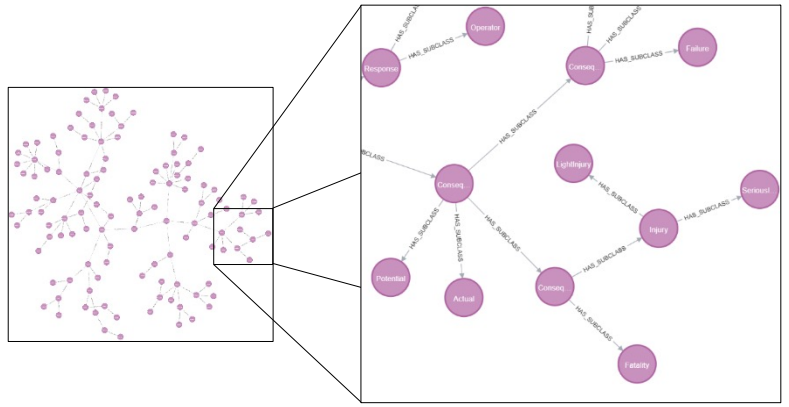
Not integrated into the work process and without support



Black-box experience
→ lack of trust

Keeping it up-to-date and relevant over time

Insights from Havtil's Incident database
from unstructured to contextualized data



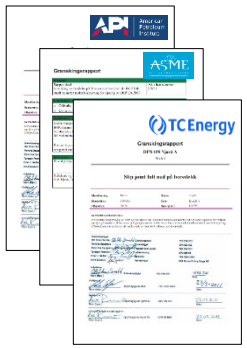
Ontology

Data in PDF

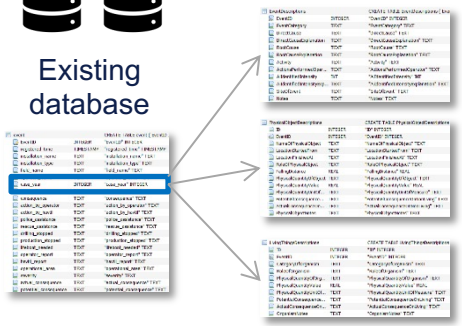
PDF text-extraction

Semantic analysis og data structuring

Knowledge integration and improvement

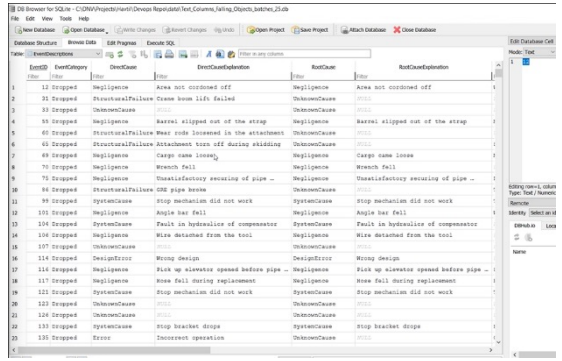


Existing database



Graf database

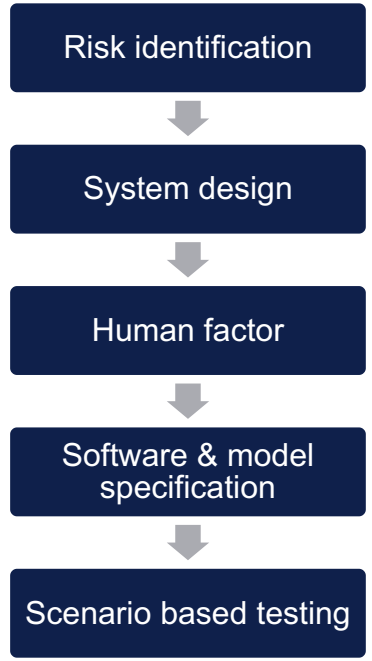
Machine-readable, contextualized and connected data, digital requirements



Unstructured data → Structured data → Connected & contextualized data → Insights

Risk assessment

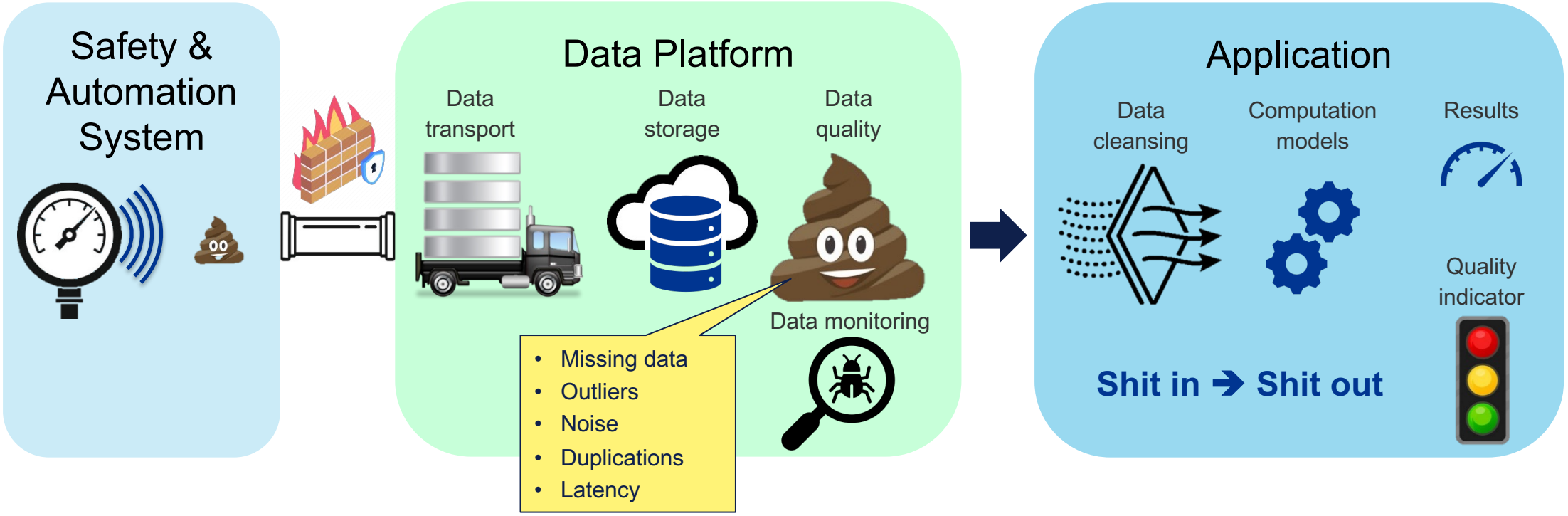
Severity	Consequence Categories			Detectability and Controllability			
				1	2	3	4
	People	Environment	Delay / Downtime	Situation can be **reliably** managed	Situation can **probably** be managed	Situation managed with **uncertainty**	Managing situation is **unreliable**
1	No or superficial injuries	Slight effect on environment, < 1 BBL	< 2 hours	L	L	L	M
2	Slight injury, a few lost workdays	Minor effect Non-compliance. < 5 BBL	< 1 day	L	L	M	H
3	Major injury, long term absence	Localized effect Spill response required. < 50 BBL	1 - 10 days	L	M	H	H
4	Single fatility or permanent disability	Major Effect Significant spill response, <100 BBL	10 - 60 days	M	H	H	V
5	Multiple fatalities	Massive effect Damage over large area, >100 BBL	> 60 days	M	H	V	V



Detectability & Controllability matrix ➔ Likelihood

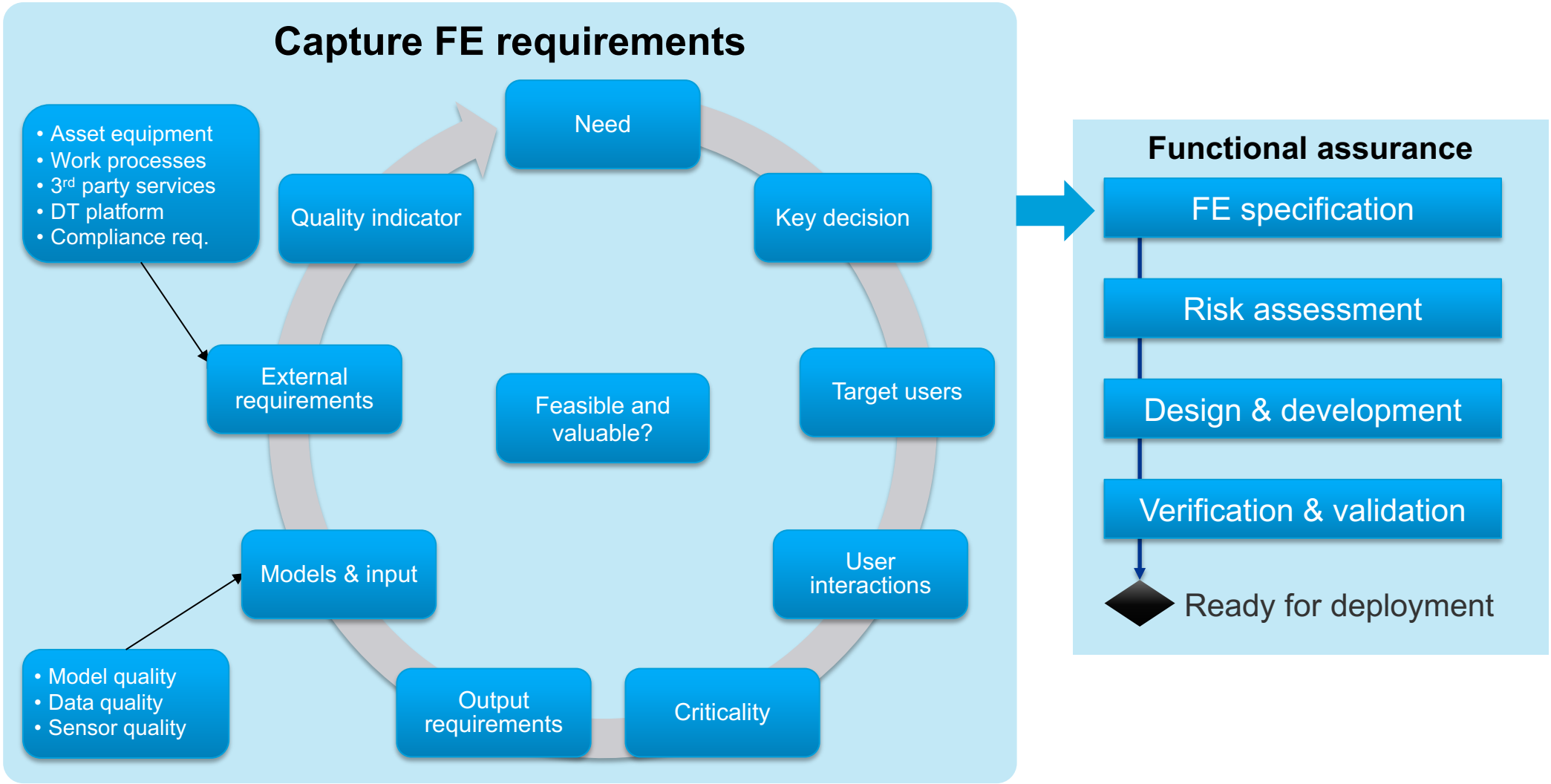
Detectability:		Controllability: M1 Controllable (>99%)	M2 Manageable (90%-99%)	M3 Difficult (70%-90%)	M4 Unreliable (<70%)
D1 Reliable	Reliable detection (>99%)	1	2	3	4
D2 Probable	Probable detection (90%-99%)	2	3	3	4
D3 Uncertain	Uncertain detection (70%-90%)	3	3	4	4
D4 Unreliable	Unreliable detection (<70%)	4	4	4	4

Data quality – someone else's responsibility?



- Can not create a trustworthy digital system without managing data quality
- AI training on uncleansed data ➔ over-fitting, unreliable, and increased maintenance cost

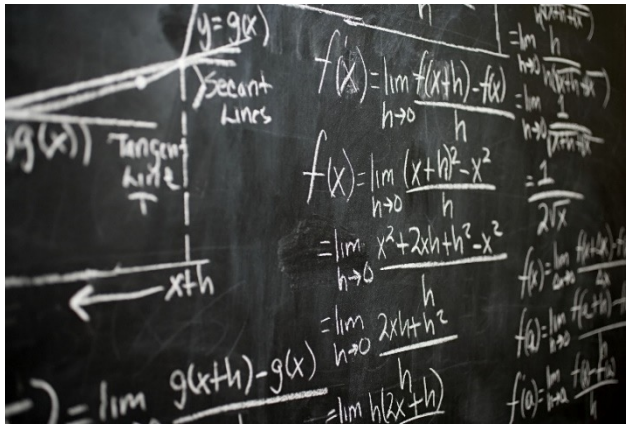
Functional Elements – Development process



Building trustworthy AI

Bringing all knowledge into AI – not just data

Science-guided AI



- Combine physical-models with data-driven models (“hybrid-models”)
- Include physical constraints into AI
- Use synthetic data for safety-critical scenarios with little data

Uncertainty-aware AI

- AI that knows when it is uncertain
- Avoid confident mistakes
- Tells you when you cannot trust it

Causal AI

- Understands cause-and-effect
- Distinguishing correlations and causation
- Distinguishing observation from intervention