

AI Application in High-Risk Systems

State-of-the-Art and current Challenges

Omar Lone - ZHAW Institute of Embedded Systems

1

Traditional OT

Determinism, traceability, verification

2

AI in OT

Black-box behaviour, non-determinism

3

Three Industrial AI Models

PINNs, neuromorphic, edge ViTs, multimodal PdM

4

EU AI Act and CRA

High-risk duties for OT

5

Use-Case and Horizon

Real deployments today

6

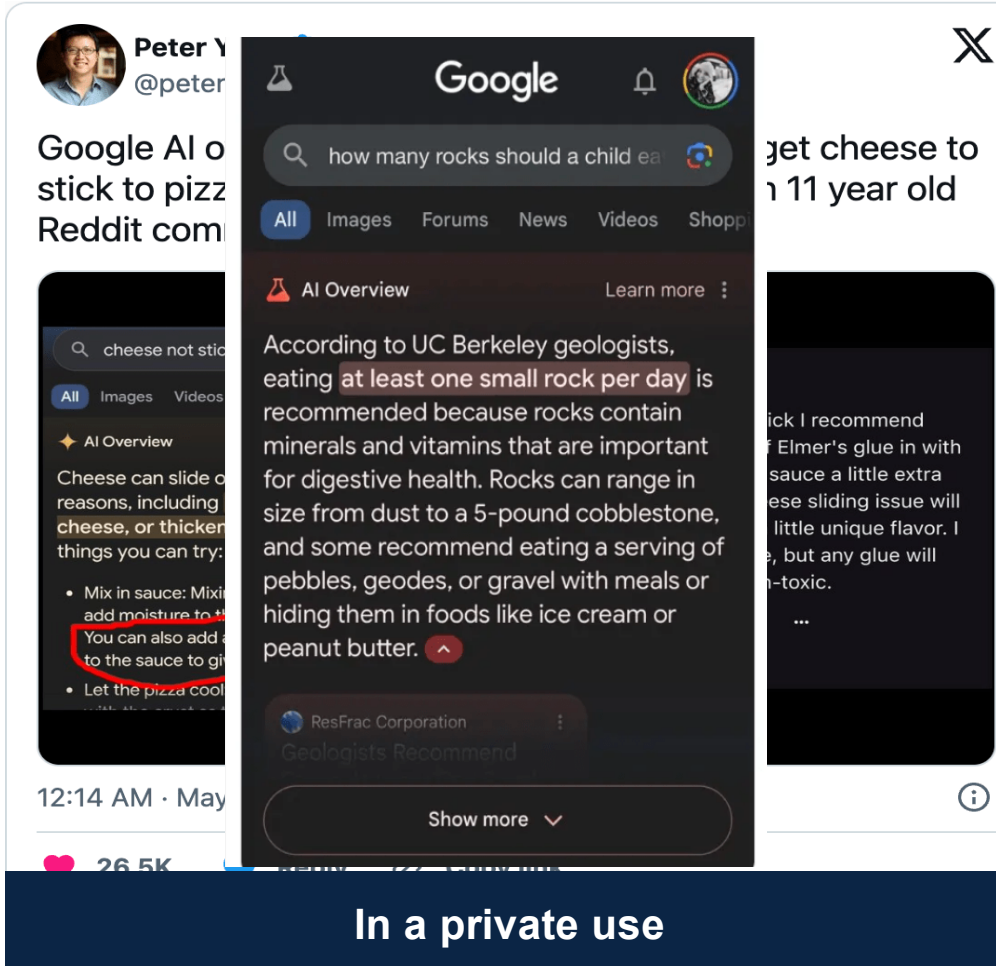
Takeaways

AI Safety Errors: 2 Scenarios

AGENDA

OT Foundations

AI in OT
Models
Regulation
Horizon
Takeaways



Peter Y (@peter)

Google AI o
stick to pizz
Reddit com

Google search: how many rocks should a child eat

AI Overview

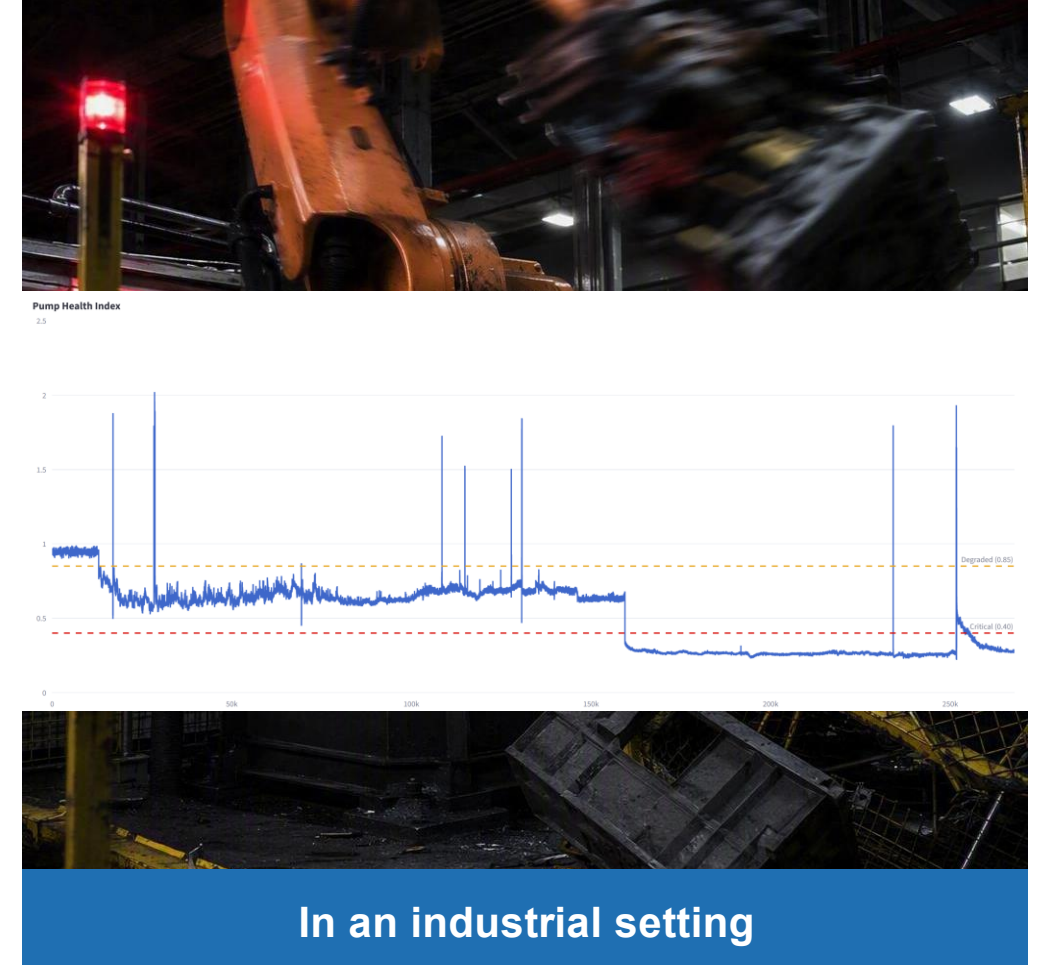
According to UC Berkeley geologists, eating at least one small rock per day is recommended because rocks contain minerals and vitamins that are important for digestive health. Rocks can range in size from dust to a 5-pound cobblestone, and some recommend eating a serving of pebbles, geodes, or gravel with meals or hiding them in foods like ice cream or peanut butter.

ResFrac Corporation
Geologists Recommend

12:14 AM · May

26.5K

In a private use



Pump Health Index

2.5

2

1.5

1

0.5

0

Degraded (0.85)

Critical (0.40)

In an industrial setting

Traditional OT (Operational Technology)

AGENDA

OT Foundations

AI in OT
Models
Regulation
Horizon
Takeaways

◆ Purpose and priorities

Safe, deterministic control
Uptime, safety, hard real time

🛡 Rule-based bounds

Beyond threshold -> defaults to a fail-safe state

⚙ No change at runtime

No dynamic code loading
Every change is reverified

Codified in standards

IEC 61508 Functional Safety

IEC 61511 Process Plants

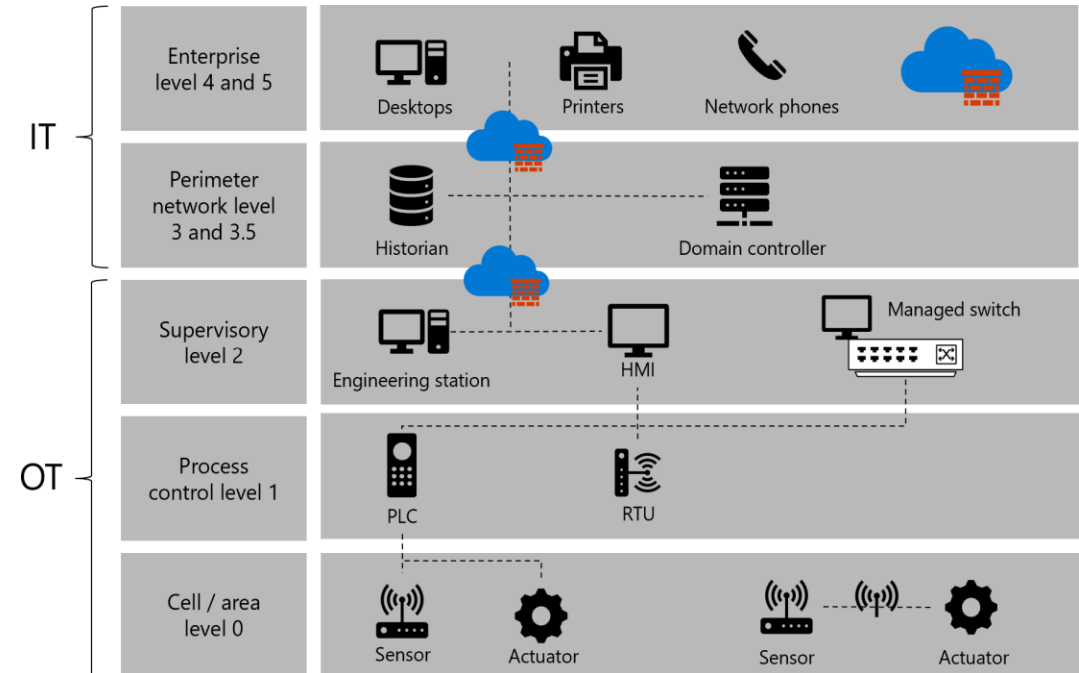
IEC 62443 Cybersecurity

ISO 13849 Machinery

IEC 61131-3 Programming Languages

ISA-95 Enterprise Control

Purdue for OT and IT separation



<https://learn.microsoft.com/en-us/azure/defender-for-iot/organizations/best-practices/understand-network-architecture>

AGENDA

OT Foundations

AI in OT
Models
Regulation
Horizon
Takeaways



Deterministic Execution

Cycle: read inputs, run logic, write outputs

Process image freezes inputs
-> no race conditions

Watchdog forces a safe state on overrun.

Low-millisecond cycles guaranteed by design



Transparent by Construction

Finite set of explicit rules -> every output traces to its code line

Written in IEC 61131-3 languages, from ladder to structured text, readable by any qualified engineer



Verifiable before release

Bounded logic and finite state space:

Exhaustively tested and formally checked against safety requirements

What the safety standards assume: IEC 61508 and ISO 13849 certify a function only when its worst-case behaviour is bounded, repeatable and provable

AGENDA

OT Foundations

- AI in OT
- Models
- Regulation
- Horizon
- Takeaways

These properties pay off the day something goes wrong

Reproducible

A fault reproduces on demand, on the bench

Bounded search

The search for a cause is finite, not open-ended

Repeatable test

Hardware-in-the-loop and static analysis pin it down

Linear causality

Every state change is logged in order

Evidence that holds

Immutable logs stand up to an auditor

Safety claims survive

SIL under IEC 61508 or PL under ISO 13849 stays claimable.

Provable fix — The investigation ends at the exact line of logic that set the output.

Neural Networks: Weights, Probabilities and Opacity

AGENDA

- OT Foundations
- AI in OT**
- Models
- Regulation
- Horizon
- Takeaways

Weights, not rules

Weighted sums tuned by gradient descent, not rules an engineer can read

A probability, not an answer

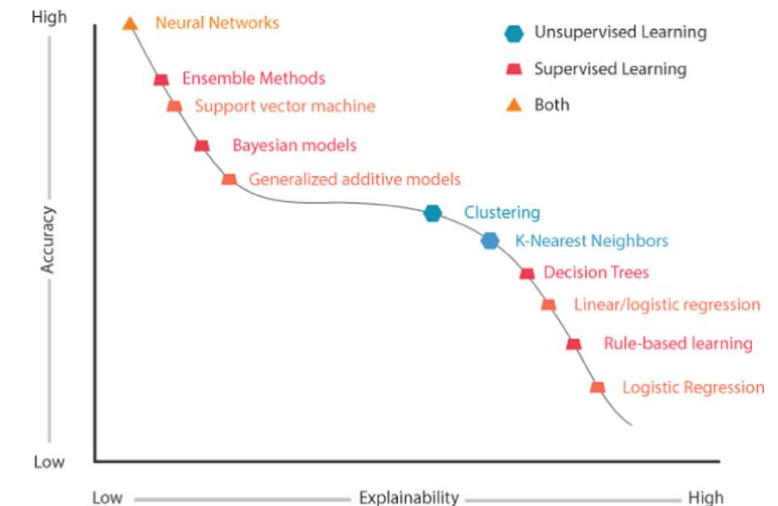
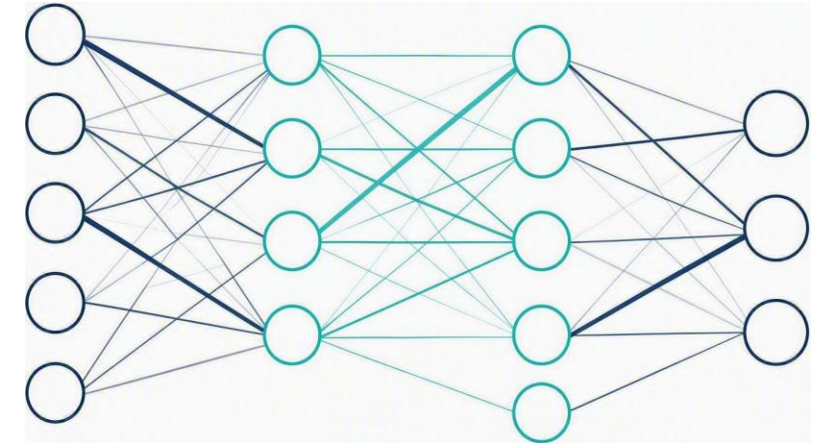
A softmax score of 0.98 is a ranking, not a guarantee

Non-deterministic in practice

Retraining, GPU scheduling and library versions shift the result

The trade-off is structural

The architectures that perform best explain themselves least



XAI: Explainability Helps, but It Is Not Assurance

AGENDA

- OT Foundations
- AI in OT**
- Models
- Regulation
- Horizon
- Takeaways

LIME and SHAP

Local explanation methods that attribute one prediction to input features

Grad-CAM

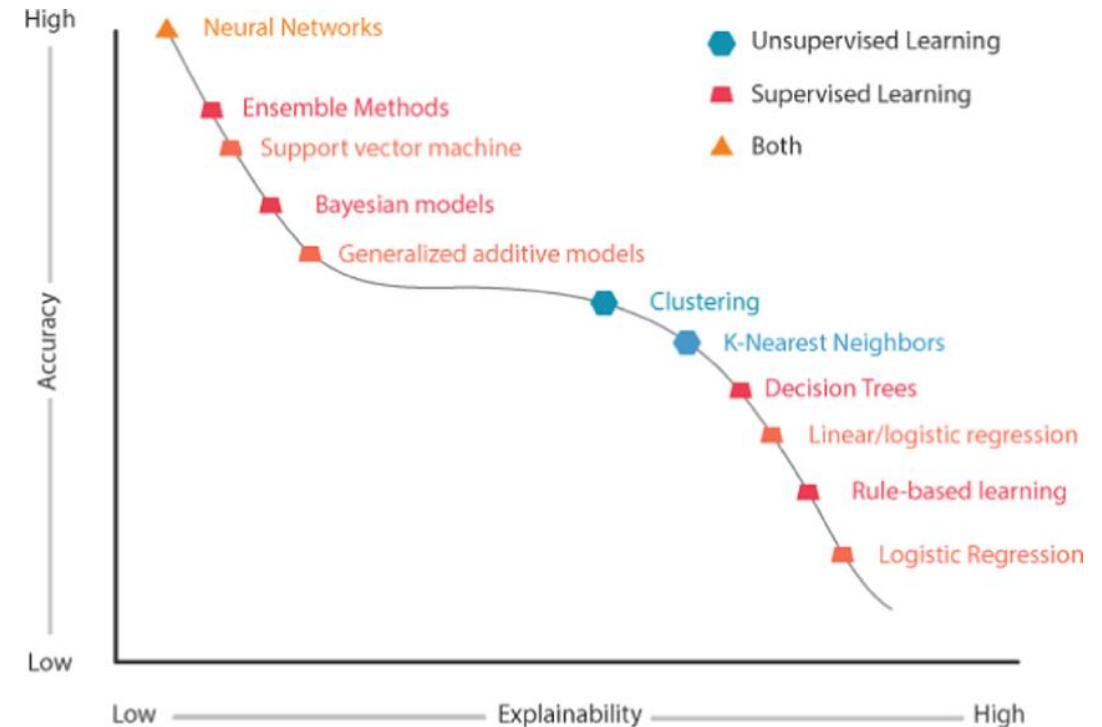
Highlights the image region behind a call

Surrogate trees

Rules an operator can actually read

Explanation is not assurance

Only approximates the model
Methods disagree
After the fact — never a safety function



https://www.researchgate.net/figure/Accuracy-versus-explainability-in-Artificial-Intelligence-models_fig3_372530108

The Collision: Deterministic Control Meets a Probabilistic Black Box

AGENDA

OT Foundations

AI in OT

Models

Regulation

Horizon

Takeaways

	Traditional OT	Neural Networks
Logic	An explicit Boolean rule	Millions of opaque weights
Output	A repeatable function	A probability distribution
Timing	Fixed worst-case timing	Variable inference latency
Failure	A defined fail-safe state	A best-matching guess
Verification	Formal proof	Empirical test sets

Key takeaway — deterministic and probabilistic systems are verified in different ways.

Why AI Is Pushing Onto the Plant Floor

Drivers

AGENDA

OT Foundations

AI in OT

Models

Regulation

Horizon

Takeaways

**Multivariate
control**

**Visual
inspection**

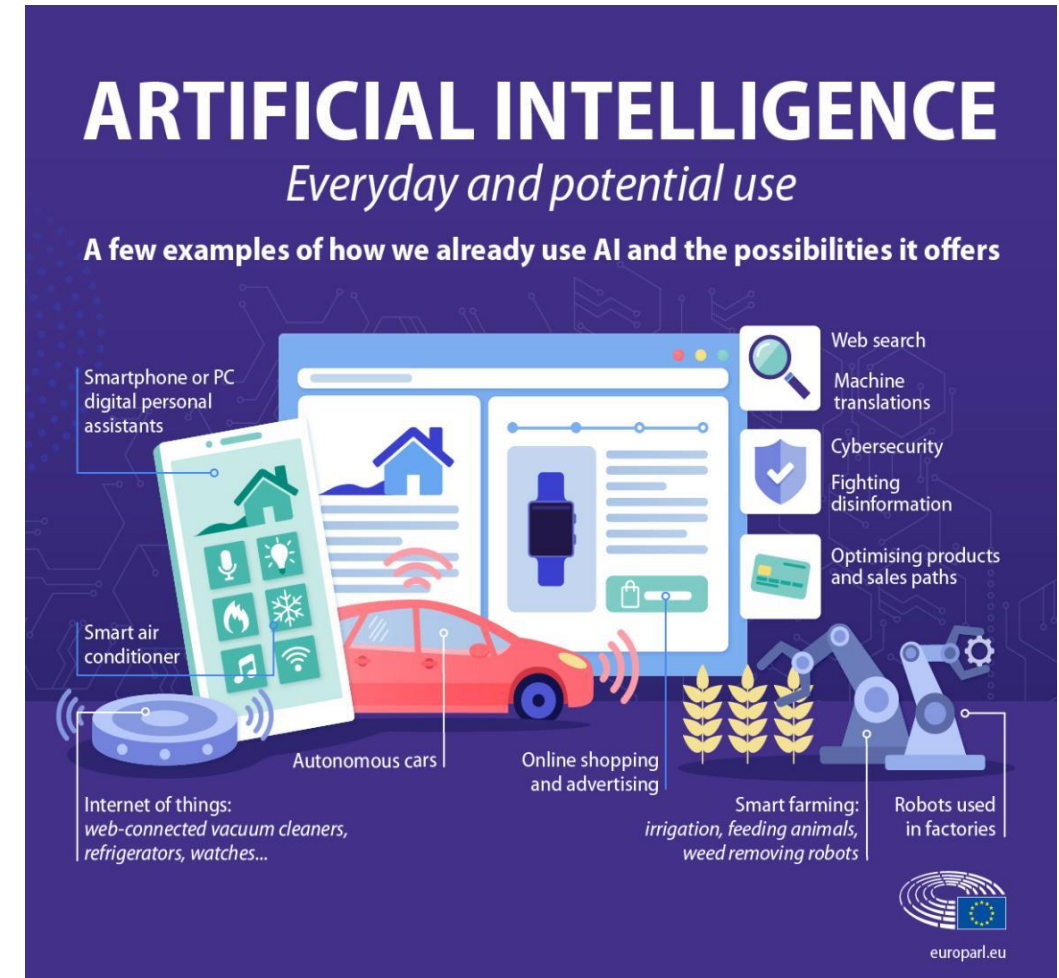
**Predictive
upkeep**

Hand-written logic cannot capture non-linear behaviour

AI is already in use but not controlling actuators

The best-performing models explain themselves least

**OT needs the most defensible model,
not the most accurate**



<https://www.europarl.europa.eu/topics/en/article/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used>

State of the Art: AI as an Advisor, Not a Controller

AGENDA

OT Foundations

AI in OT

Models

Regulation

Horizon

Takeaways



Deterministic clash

A PLC must guarantee a bounded, repeatable response
A Neural Network model cannot



No neural network in the control path

No certified deployment puts a neural network in the control path



Advisory role instead

It reads process data and proposes a state, setpoint, class or alert



A human or a rule decides

The operator confirms, or the output enters as a bounded parameter

Containing Non-Determinism: Three Architectural Patterns

AGENDA

OT Foundations

AI in OT

Models

Regulation

Horizon

Takeaways

1

Advisor

Runs out of loop and proposes an action

Why it works: a human or a rule disposes
simplest pattern to justify.

2

Simplex enclave

AI sets the setpoint

A deterministic controller clips it to hard bounds

Guardrail: checked against physical limits, rate limits and interlocks

3

Fallback ensemble

Reverts to certified PID or state-machine logic when plausibility drops

Design point: the fallback must run the process, not just stop it

No neural network drives an actuator without a deterministic layer between them

AGENDA

OT Foundations

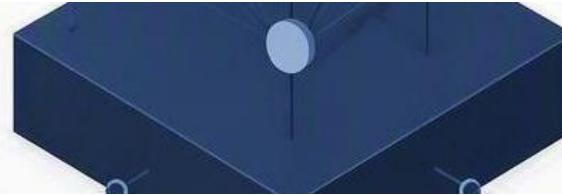
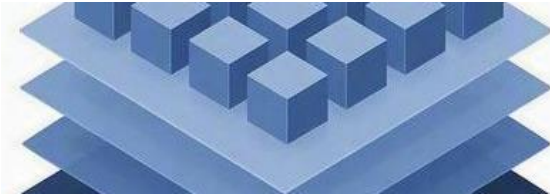
AI in OT

Models

Regulation

Horizon

Takeaways



Standard neural network

Optimizes fit alone: ignores conservation laws, needs large labelled sets

OT adds three constraints

Physical truth, low latency at low power, mixed sensors

PINNs

Physics equations enter the training loss

Answers: physical truth

Edge vision

Edge Vision Transformers run in the machine

Answers: low latency at low power

Multimodal PdM

Fuses vibration, acoustics and thermal data

Answers: heterogeneous sensors

Model 1: Physics-Informed Neural Networks (PINNs)

AGENDA

OT Foundations

AI in OT

Models

Regulation

Horizon

Takeaways



Definition

Trained to fit data while minimizing the residual of the governing equations



Loss

Total loss = data loss + weighted physics loss, so violations are penalised



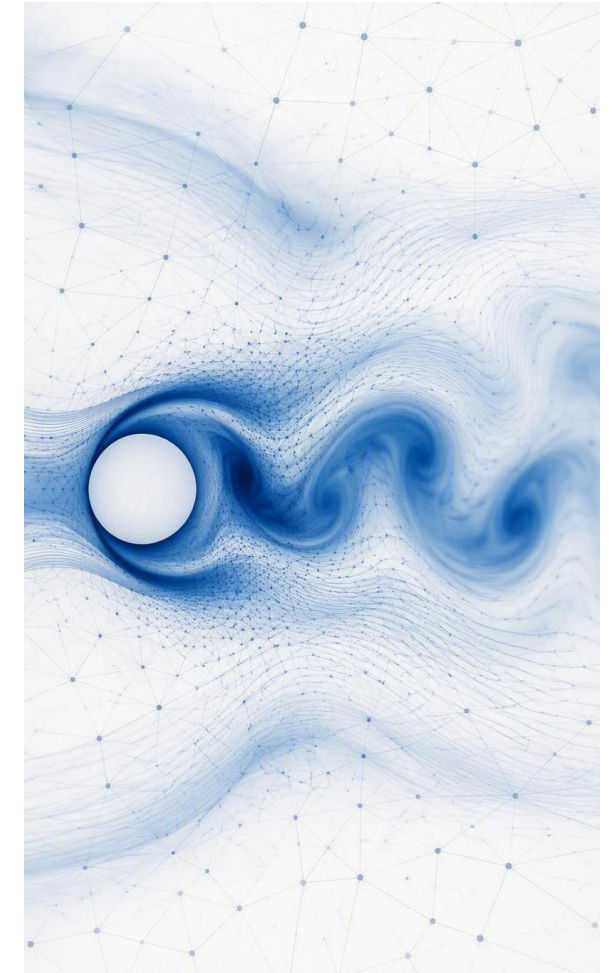
Typical uses

Heat in casting, flow in pipelines, fatigue in rotating equipment



Sparse data and inverse problems

Physics regularizes unobserved regions



PINNs versus a Standard Neural Network

AGENDA

OT Foundations

AI in OT

Models

Regulation

Horizon

Takeaways

Standard neural network

Data

Massive labelled sets

Extrapolation

Unreliable outside training data

Consistency

Can predict a negative mass

Physics-Informed Neural Network

Data

Sparse data plus a known equation

Extrapolation

Behaviour anchored to physics

Consistency

Penalised for a negative mass

Honest caveat: a soft penalty, not a hard guarantee; monitor residuals at runtime

Cost: slower training; stiff or multi-scale equations are hard to converge

Verdict: strong for thermal, fluid and structural loops, not a drop-in for perception

Model 2: Neuromorphic Vision and Edge Vision Transformers

AGENDA

OT Foundations

AI in OT

Models

Regulation

Horizon

Takeaways

Frame Cameras Today

30–60 fps: fast motion blurs,
network floods

Event Cameras

Async per-pixel changes,
microsecond timing

Activity-Scaled Data

Bandwidth scales with activity,
not pixels

Edge ViTs

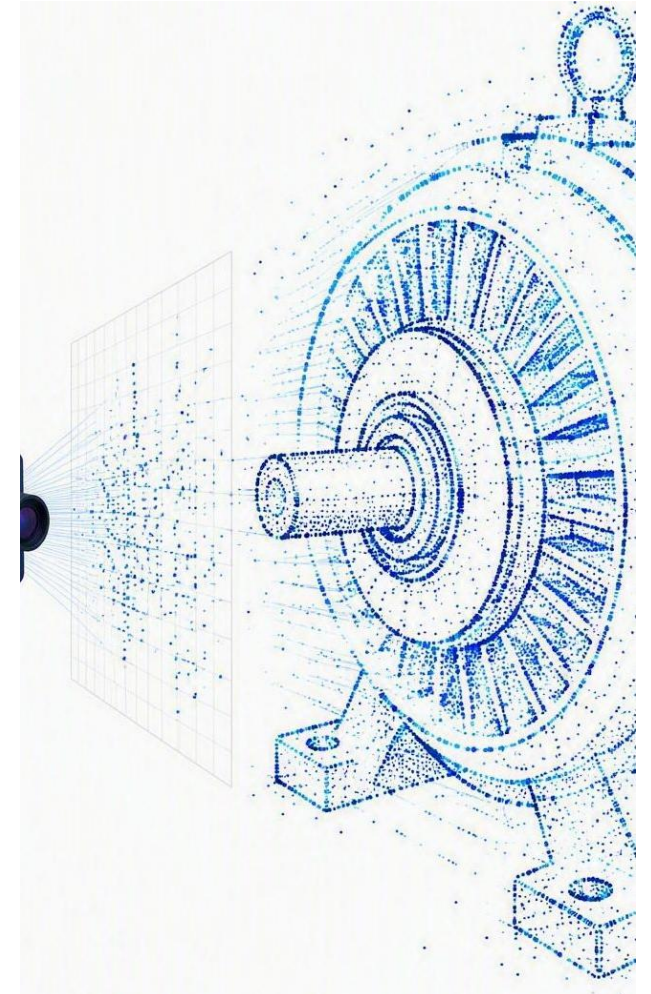
Quantized and pruned; adds
global context

Runs On-Machine

NPU, FPGAs or
neuromorphic silicon
No cloud

Together

Sub-frame reaction, decided
at the sensor



Neuromorphic and Edge ViT versus a Standard CNN (Convolutional Neural Network)

AGENDA

OT Foundations

AI in OT

Models

Regulation

Horizon

Takeaways

Neuromorphic and Edge ViT	Standard CNN
Sensing Asynchronous events, microsecond timing	Fixed frame rate, motion blur
Latency Sub-millisecond event response	Tens of milliseconds per frame
Power Activity-driven, milliwatt range	Continuous dense computation
Context Global self-attention across the part	Local receptive fields
Honest caveat Event toolchains are thin; quantisation can cost accuracy.	

Model 3: Multimodal Predictive Maintenance

AGENDA

OT Foundations

AI in OT

Models

Regulation

Horizon

Takeaways



Single-sensor blind spots

Vibration misses winding faults



Inputs fused

Accelerometers, acoustic, current, thermal, load



Cross-attention layers

Aligns streams at different sampling rates



Output

Degradation state and remaining useful life



Value

Bearing, insulation and gear faults weeks early



<https://www.yeschamp.com/news/The-Ultimate-Guide-to-Horizontal-Centrifugal-Pumps-Efficiency-Applications-and-Selection-Tips.html>

Multimodal Fusion versus a Single-Stream Model

AGENDA

OT Foundations

AI in OT

Models

Regulation

Horizon

Takeaways

Multimodal Fusion	Single-Stream Model
Coverage Mechanical, electrical and thermal modes	One failure family only
False alarms Cross-checks between modalities	High sensitivity to load and speed
Remaining useful life Adapts to the operating regime	Degrades under changing duty cycles
Cost Time sync, hardware, integration work	A simple sensor interface
Honest caveat Fusion needs run-to-failure data most plants lack	

Comparative Summary: Where Each Model Belongs

AGENDA

OT Foundations

AI in OT

Models

Regulation

Horizon

Takeaways

General deep networks

Strong perception, weak physical guarantees;
keep out of the control loop

PINNs

Thermal, fluid and structural processes:
equations known, data scarce

Neuromorphic and edge ViTs

High-speed inspection and reaction under strict
power and latency limits

Multimodal maintenance

Asset health and planning: a wrong call costs a
work order, not a life

The pattern More physics or redundancy encoded, less dependence on the model

None is exempt: the EU AI Act judges every architecture here

The EU AI Act Arrives in the Machine Hall

AGENDA

OT Foundations

AI in OT

Models

Regulation

Horizon

Takeaways

- **1 August 2024** — In force, staged
- **February 2025** — Prohibitions, AI literacy
- **August 2025** — General-purpose duties
- **27 July 2026** — Digital Omnibus defers
- **December 2027** — Annex III deadline
- **August 2028** — Embedded AI deadline

Four risk tiers

Banned, high risk, transparency, minimal

Penalties

EUR 35 million or 7% turnover (prohibited); EUR 15 million or 3% (other)



<https://artificialintelligenceact.eu/ai-act-explorer/>

Why Industrial AI Usually Lands in the High-Risk Class

AGENDA

OT Foundations

AI in OT

Models

■ **Regulation**

Horizon

Takeaways

Annex I Section A

AI as a safety component of a product already under EU harmonisation law, machinery and medical devices included

Also in scope

Machinery Regulation (EU) 2023/1230, applicable from January 2027, plus lift, pressure and ATEX equipment

Annex I Section B

Partial carve-out: automotive, aviation, rail and marine keep sectoral frameworks that must absorb AI Act rules

Annex III

Critical infrastructure: safety components in electricity, water, gas and heating supply

Typical cases

Cobot force adaptation, AMR navigation, vision driving an interlock

Consequence

An unvalidated black box that moves an actuator is not a compliant product in the EU

What High Risk Actually Demands: Articles 8 to 15

AGENDA

OT Foundations

AI in OT

Models

■ Regulation

Horizon

Takeaways

- 1 Article 9, risk management** Continuous, documented cycle across the whole lifecycle
- 2 Article 10, data governance** Relevant, representative, complete training and test data
- 3 Articles 11 and 12** Technical documentation before market placement; automatic logging in operation
- 4 Article 13, transparency** Instructions stating capability, limits and expected accuracy
- 5 Article 14, human oversight** Operator must be able to interpret, intervene and stop the system
- 6 Article 15** Accuracy, robustness, cybersecurity; resilience to drift and adversarial attack

You Do Not Need a New Framework: Map It Onto Standards

AGENDA

OT Foundations

AI in OT

Models

Regulation

Horizon

Takeaways

Safety and security you already run

IEC 61508 and ISO 13849

Keep functional safety intact through SIL and Performance Level

IEC 62443

Secures the pipeline, from sensor ingest to the edge inference node

SOTIF, ISO 21448

Hazards from performance limits, not faults — exactly the AI case

AI-specific layers to add

ISO/IEC 42001:2023

AI management system: policies, roles, impact assessment, improvement

ISO/IEC TR 24029

Methods for assessing the robustness of neural networks

Harmonised AI Act standards

Still drafting in CEN-CENELEC JTC 21; none yet in the Official Journal

The Cyber Resilience Act: Security as a Product Duty

AGENDA

OT Foundations

AI in OT

Models

Regulation

Horizon

Takeaways

Regulation (EU) 2024/2847 covers every product with digital elements placed on the EU market, including PLCs, gateways and edge inference nodes



December 2024

Regulation in force

September 2026

Vulnerability reporting

December 2027

Full obligations apply

Secure by design

Secure-by-design build, software bill of materials, security updates over the support period

Bridge to Article 15

CRA compliance presumes Article 15 conformity (IEC 62443 evidence)

A one-way bridge

Cybersecurity only: saves duplicated evidence, does not merge the two assessments

Pipeline consequence

Model artefacts and containers are attack surface; IEC 62443 supplies zoning and patching

Beyond Europe: How the Rest of the Market Regulates AI

AGENDA

OT Foundations

AI in OT

Models

■ Regulation

Horizon

Takeaways

United States

No federal AI act

A deliberately
deregulatory federal
direction

The NIST AI Risk
Management
Framework is the de
facto reference.

United Kingdom

Principles-based and
sector-led

Existing regulators
such as the HSE
apply safety law to AI

No new statute

China

Most prescriptive
regime outside the
EU: algorithm and
generative model
filing, security
assessment,
mandatory labelling

Asia-Pacific

South Korea's AI
Framework Act, in
effect 22 January
2026, is the closest
cousin to the EU
model

Japan:
AI Promotion Act

Singapore tests via
AI Verify

(Implied): Build for the EU AI Act and the CRA and you clear the rest

Functional Safety and Security Under One Roof

AGENDA

OT Foundations

AI in OT

Models

■ Regulation

Horizon

Takeaways

The SIL Problem

Safety integrity assumes deterministic code with provable failure modes

Decoupling

Certified safety stays deterministic

The AI advises

SOTIF

Hazards from the model performing poorly, not breaking

Three Attack Vectors

Data poisoning, adversarial inputs, model extraction

IEC 62443

Zones and conduits isolate inference; models signed, data hashed

Security Is Safety

A poisoned model is a failed safety function

A Compliant Industrial AI Pipeline, End to End

AGENDA

OT Foundations

AI in OT

Models

■ Regulation

Horizon

Takeaways

- 1 Ingest guard** Sensor data checked for integrity and provenance
- 2 Edge inference** Runs PINN, edge ViT or multimodal fusion
- 3 Explanation layer** Scores output, flags out-of-distribution data
- 4 Deterministic cage** Hard physical bounds; verify before acting
- 5 Logging** Decisions, confidence and overrides logged for Article 12
- 6 Human oversight** Dashboard lets the operator interpret and stop

AI in the Field Today: Real Deployments

AGENDA

OT Foundations

AI in OT

Models

Regulation

Horizon

Takeaways

Cobots

Universal Robots: TUV
Rheinland certified, Performance
Level d

Machine Vision

Model classifies; a deterministic
interlock rejects

Digital Twins

Rolls-Royce engines
Digitale Schiene Deutschland

Rail

Rio Tinto AutoHaul: driverless
since 2018, 53 trains

Energy

DeepMind wind forecasts at
Google: +20 percent value

Medical

LumineticsCore (IDx-DR): FDA
authorised 2018

Mobility and Aviation

Waymo cites fleet statistics
EASA assistance only



AGENDA

OT Foundations

AI in OT

Models

Regulation

Horizon

Takeaways



Formal Verification

SMT tools: Marabou, alpha-beta-CROWN; VNN-COMP

Caging the Network

Verified envelopes, monitors and interlocks

Neuro-Symbolic

Learned perception, provable symbolic logic

Neuromorphic Silicon

BrainChip Akida at milliwatts; Intel Loihi 2 at watts

Foundation and VLA

Run inside constrained operational envelopes

Federated Learning

Models improve; telemetry stays on site

Honest read: verification scales only to modest networks, so caging is the practical answer this decade

Takeaways for OT Leaders

AGENDA

OT Foundations

AI in OT

Models

Regulation

Horizon

■ **Takeaways**

AI and Determinism

They coexist only behind a deterministic guardrail.

Model Fits Physics

PINNs for process, edge vision for speed, fusion for assets.

Stay Critical

A physics loss is a penalty, not a proof.

Compliance Is Engineering Work

Build logging, lineage and documentation in now.

Start With Classification

Which AI functions are high risk, and what evidence exists today.

Thank you.

Contact: Omar Lone, lone@zhaw.ch - ZHAW Institute of Embedded Systems