

whitebox



Interpretable

explainable



ethical

transparent



trustworthy

fair

BROKERAGE EVENT | EUREKA CLUSTERS AI 2021

Explainable AI (XAI)

Fostering User Trust in Industrial AI Applications

Benjamin Kloepper, Arzam Kotriwala



Process Industry – Users and AI Applications



Knowledge Intensive



Multiple Applications



Process Driven



Data Intensive



Decision Making

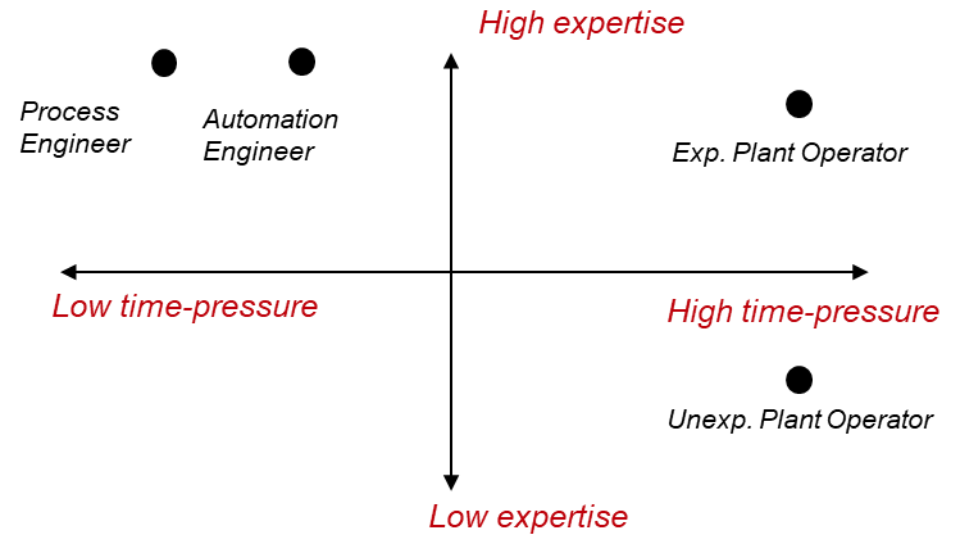


Time Pressure

Common AI Applications

Application	End Users
Process monitoring	Operator, Process engineer, Automation engineer
Fault diagnosis	Process engineer, Automation engineer, Operator, Maintenance engineer
Event prediction	Operator
Predictive maintenance	Operator, Maintenance engineer, Scheduler

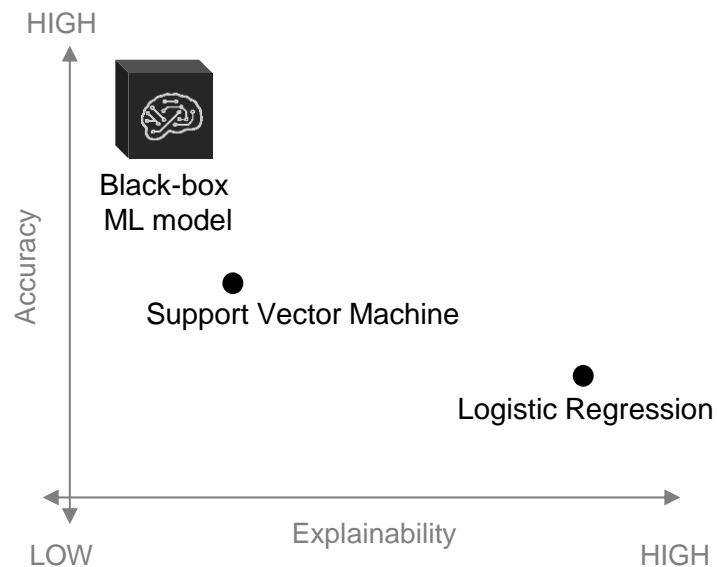
Example: Fault Diagnosis



Requirements vary across different industrial users, applications w.r.t time pressure and expertise

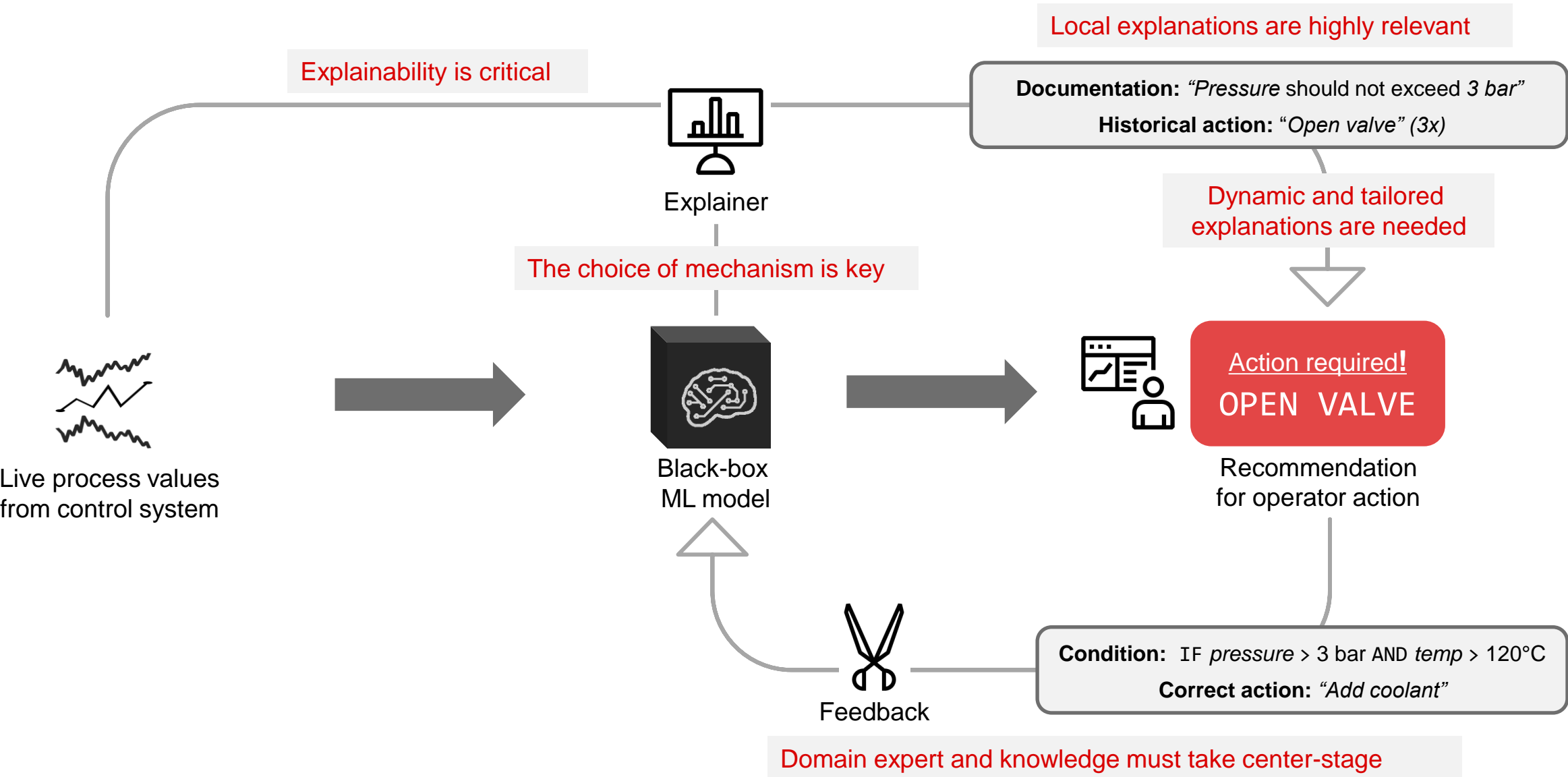
Explainable AI (XAI)

Trade-off: Accuracy VS Explainability

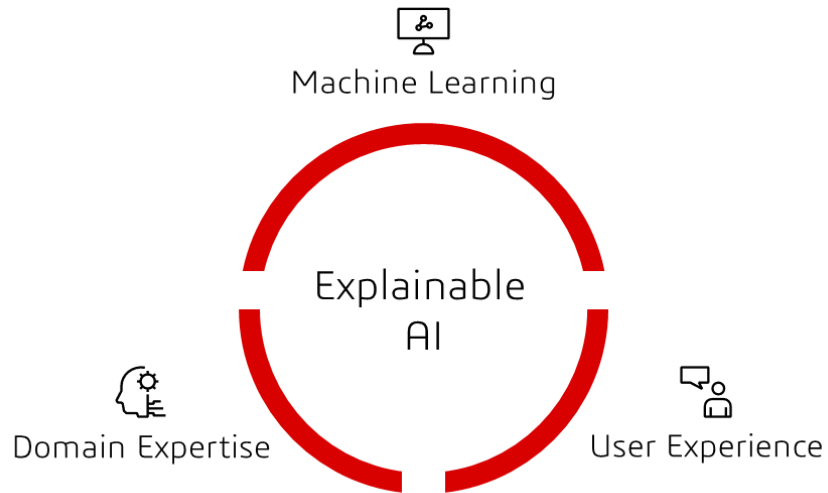


Summary of important properties of XAI approaches from literature

Property	Types
Interpretability	Intrinsic or Post-hoc
Scope	Global or Local
Mode	Interactive or Static
Supported data types	Tabular, Image, Text or Time-series
Mechanism	Attribution, Surrogate, Similarity, Prototype or Counterfactual



Proposal



User-centered approach to XAI

- How can we better understand the user and context?
- How to measure degree of explainability and understandability?
- How to collaborate with and co-design AI systems with users to build trust?

XAI methods for industrial data

- How can existing methods be adapted?
- Develop methods tailored to industrial time-series data
- How to embed domain expertise into ML methods?

A multi-disciplinary approach is needed for successful and sustainable application of XAI to process industry