FACULTY OF ENGINEERING AND ARCHITECTURE

Adaptive anomaly detection and root cause analysis by fusing semantics and machine learning

Bram Steenwinckel

Promotors: Femke Ongenae & Filip De Turck

Mentor: Anna Lisa Gentile

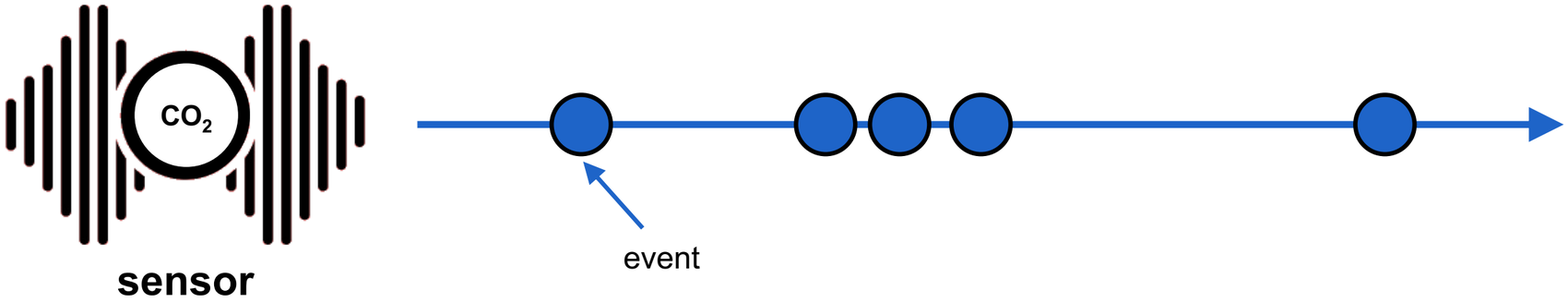
GHENT UNIVERSITY

IDLab
INTERNET & DATA LAB

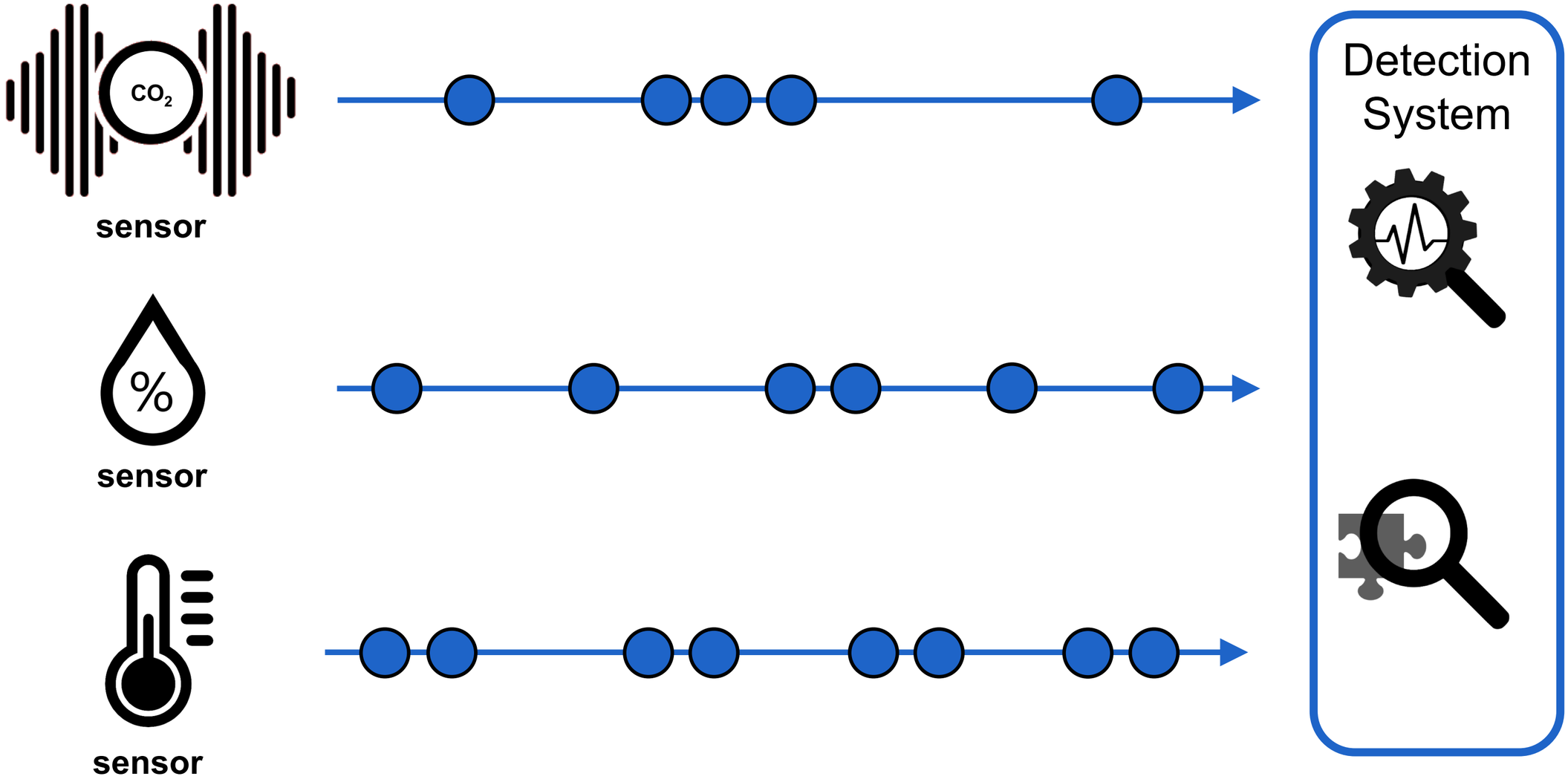
ESWC18

imec

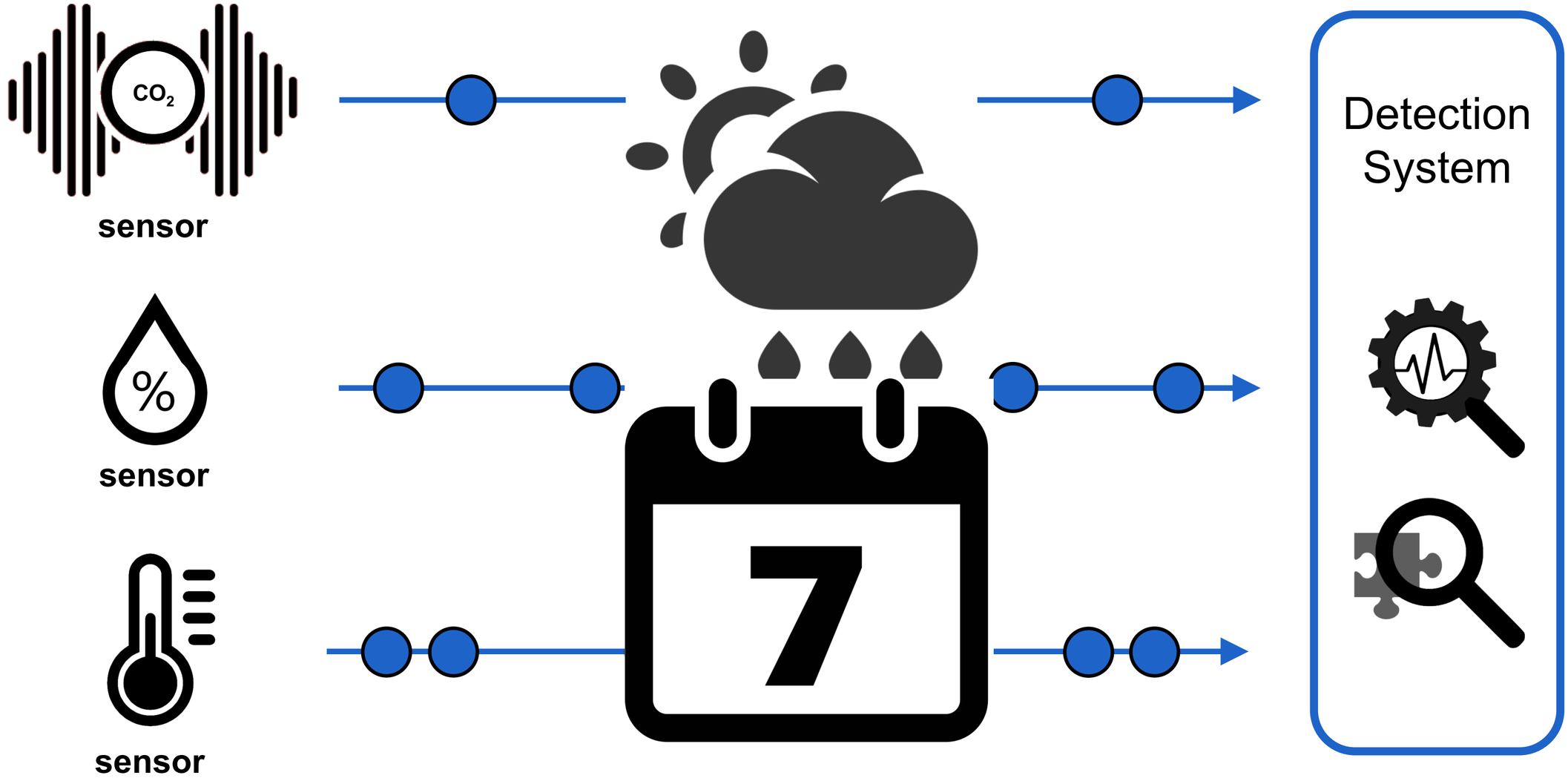
SENSOR MONITORING



SENSOR MONITORING



SENSOR MONITORING



COMMON TOOLS

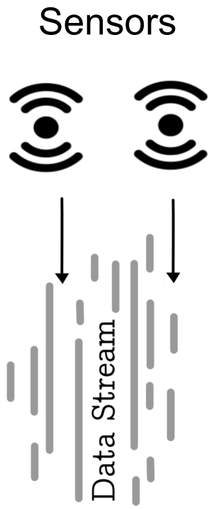
Anomaly detection

Root cause Analysis

DETECTION MECHANISMS TODAY

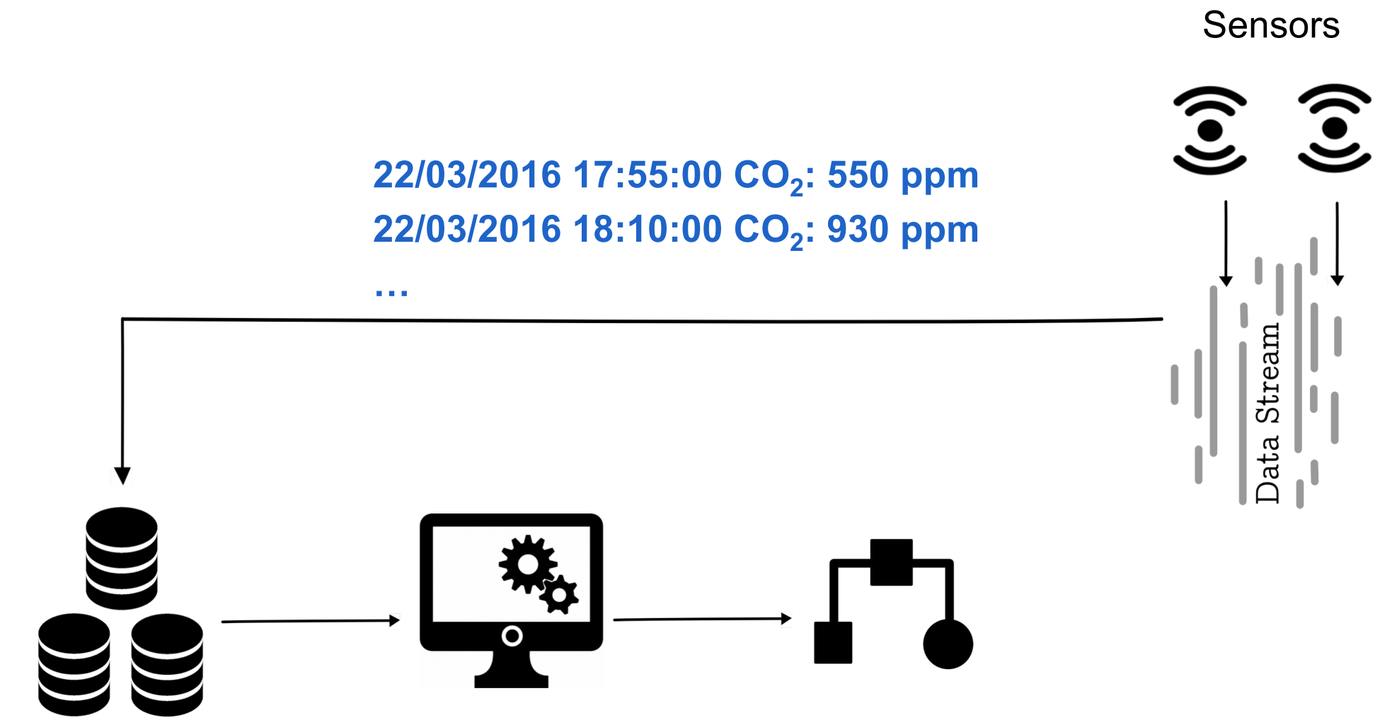
Sensors

DETECTION MECHANISMS TODAY

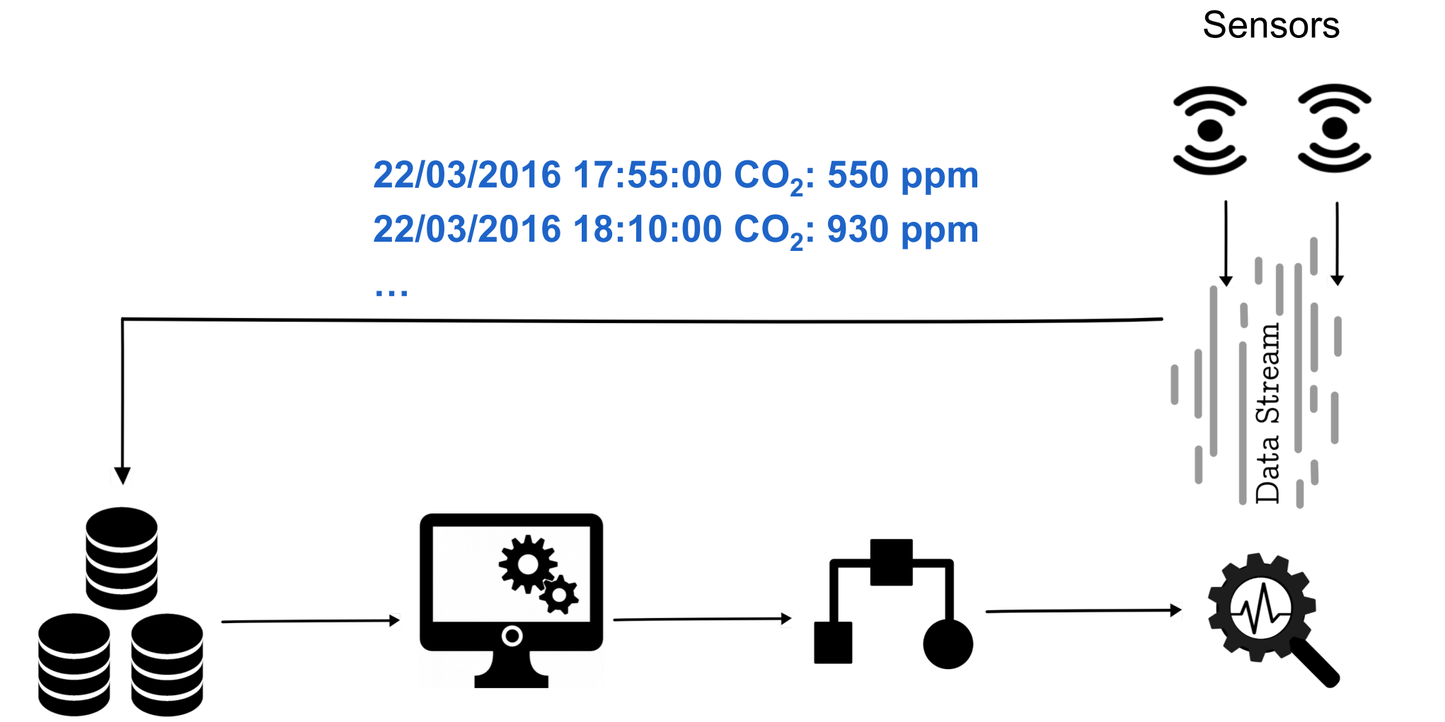


DETECTION MECHANISMS TODAY

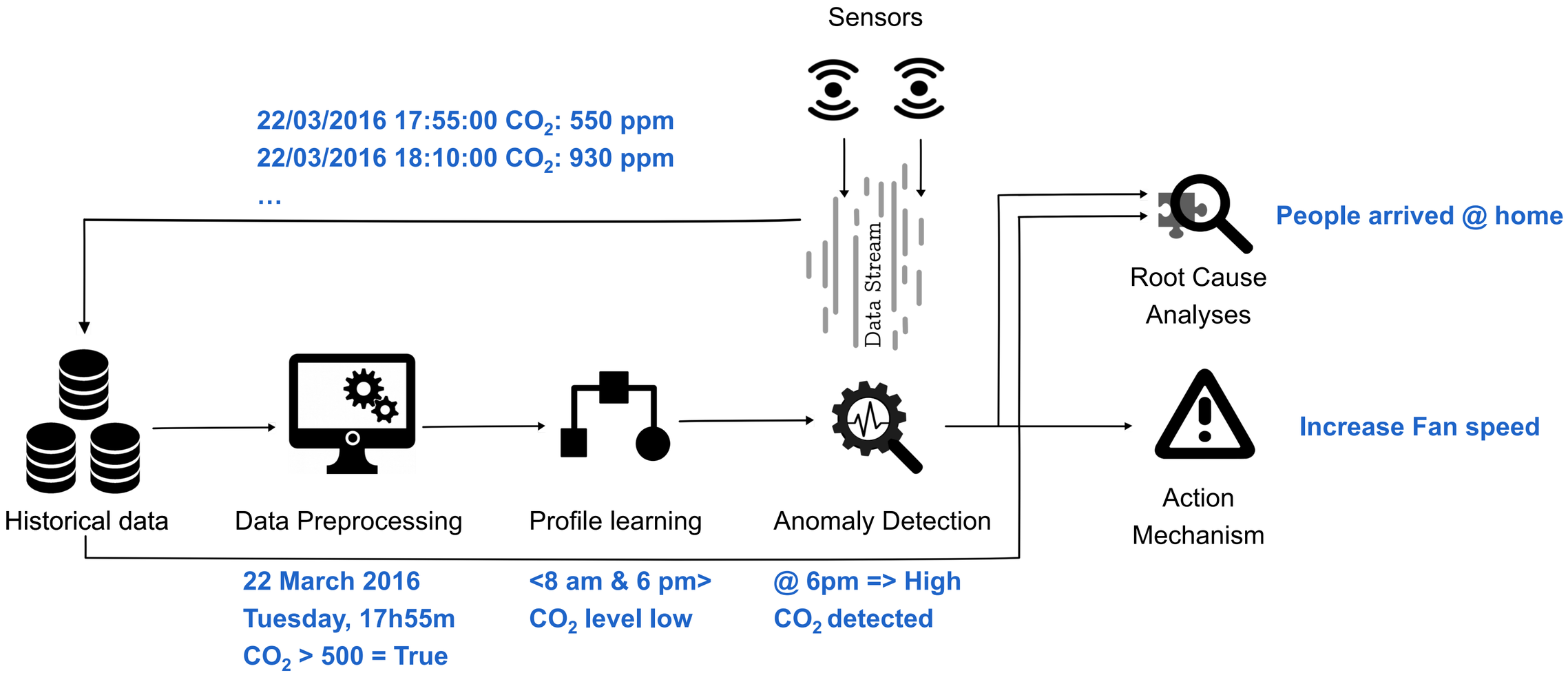
DETECTION MECHANISMS TODAY



DETECTION MECHANISMS TODAY



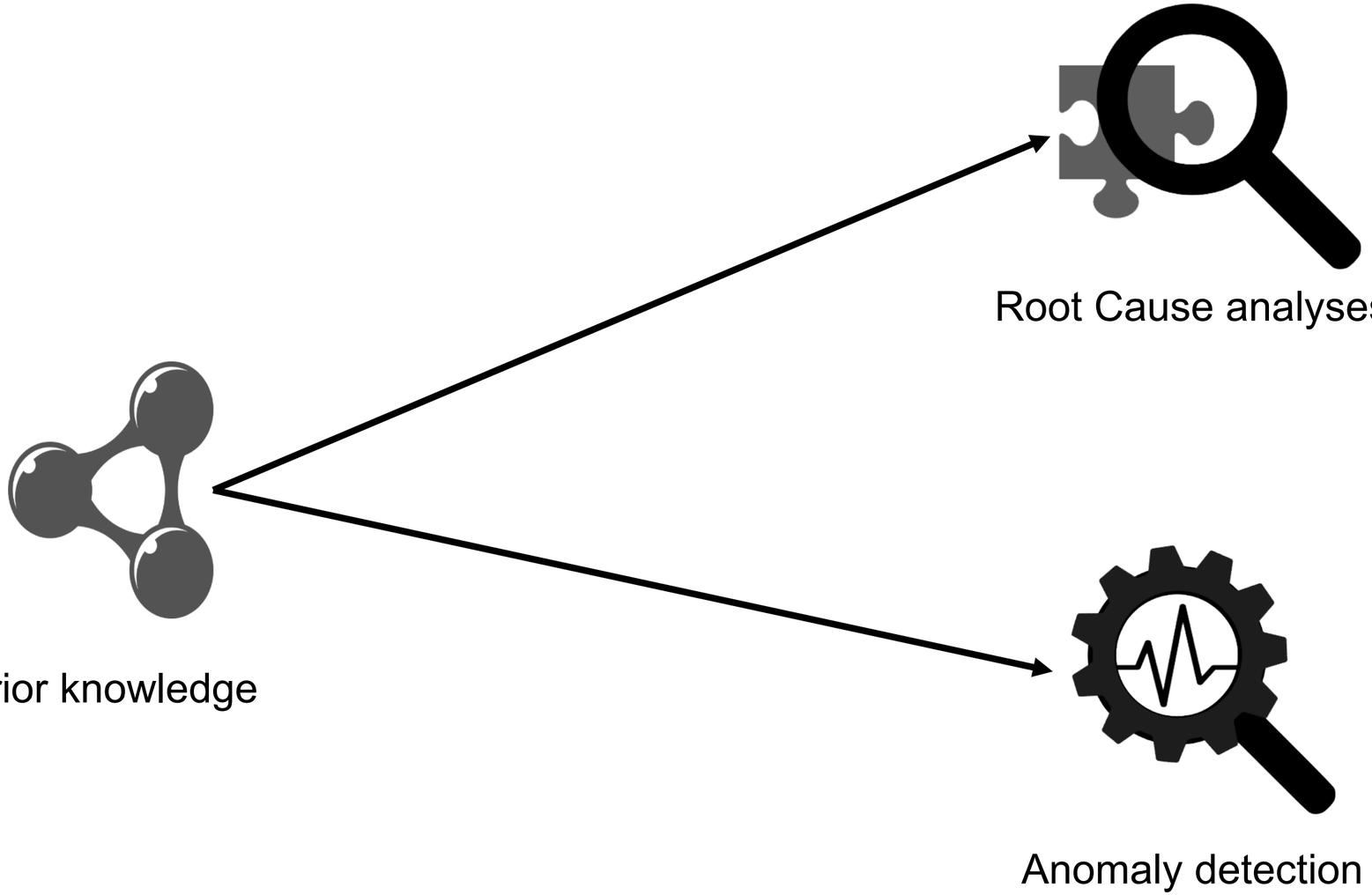
DETECTION MECHANISMS TODAY



PROBLEM

> “The AD and RCA tools of today have difficulties to adapt to changing environments without a lot of human involvement…”

GOAL?



- Reduce the human involvement
- Apply in streaming environment

- Increase detection rate for real time cases
- Reduce number of wrong detections (F1 score)

HOW TO EVALUATE?

1) RDF Datasets for classification (AM, AIFB, MUTAG)

⇒Adapted to detect the minority class(es).

2) Proof of concept case studies:

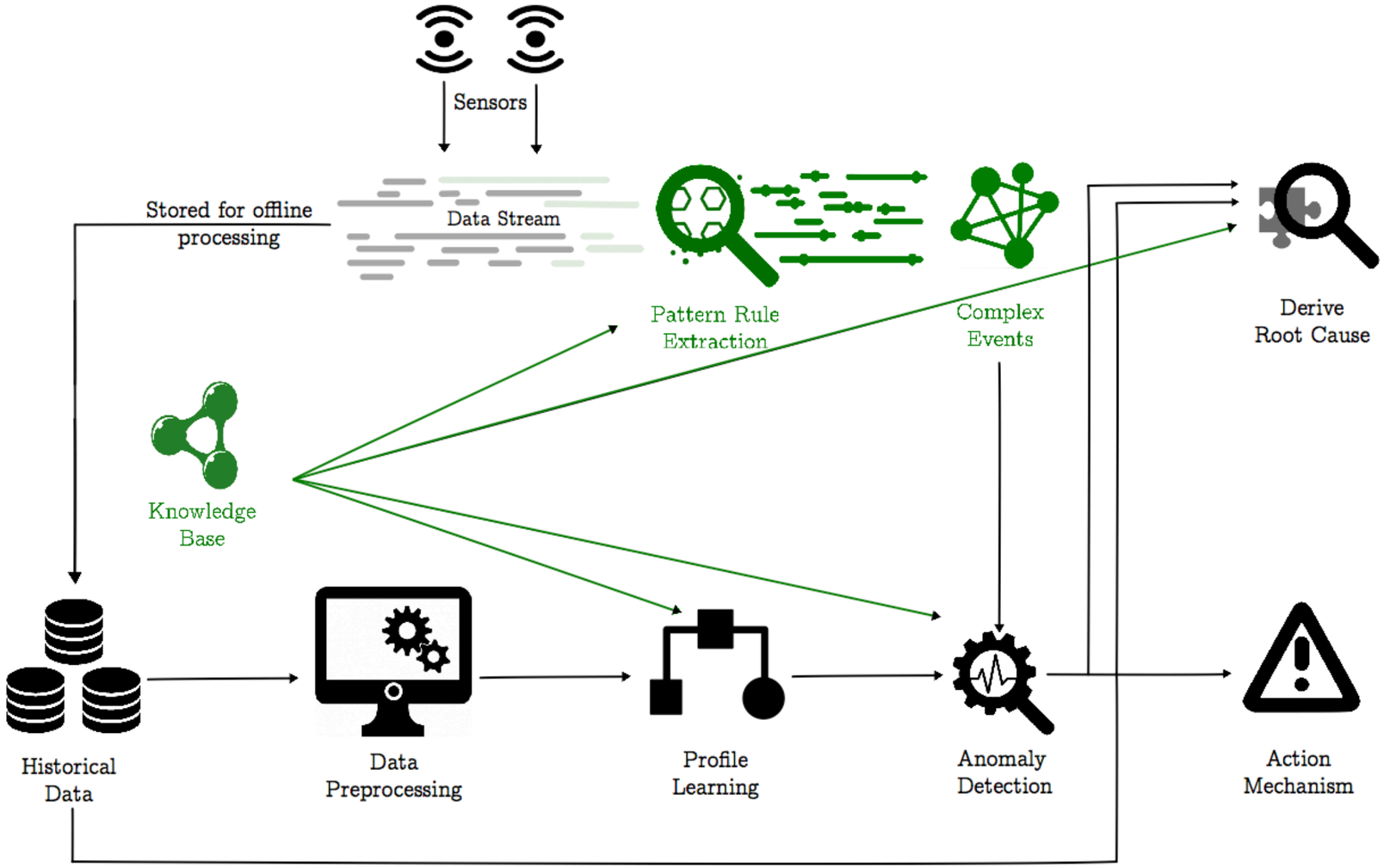
Pervasive healthcare

Predictive maintenance

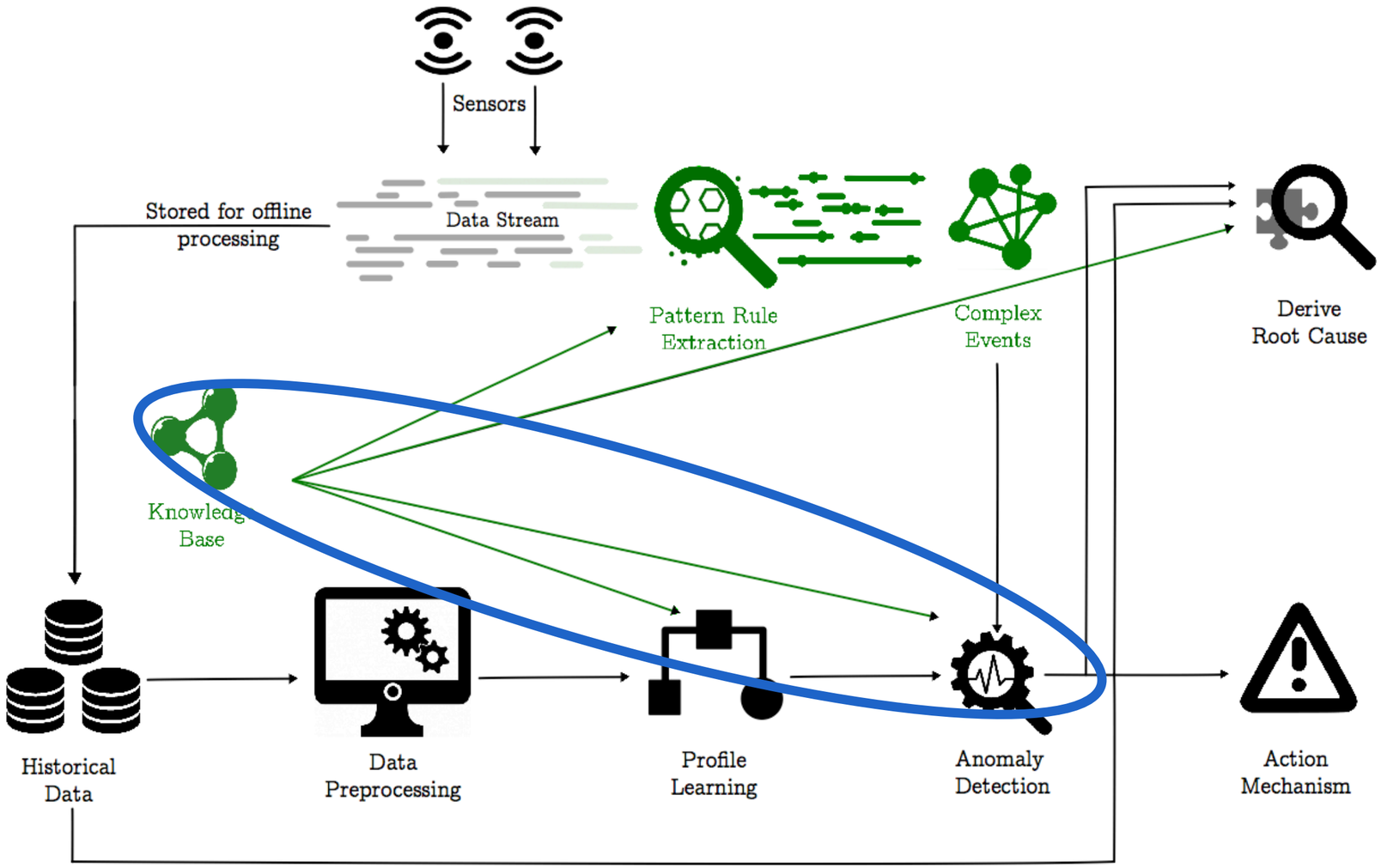
OUTLINE

1. Introduction
2. Feature importance with knowledge learning
3. Interpretable knowledge for cause detection
4. Adaptive stream detection
5. Additions & conclusion

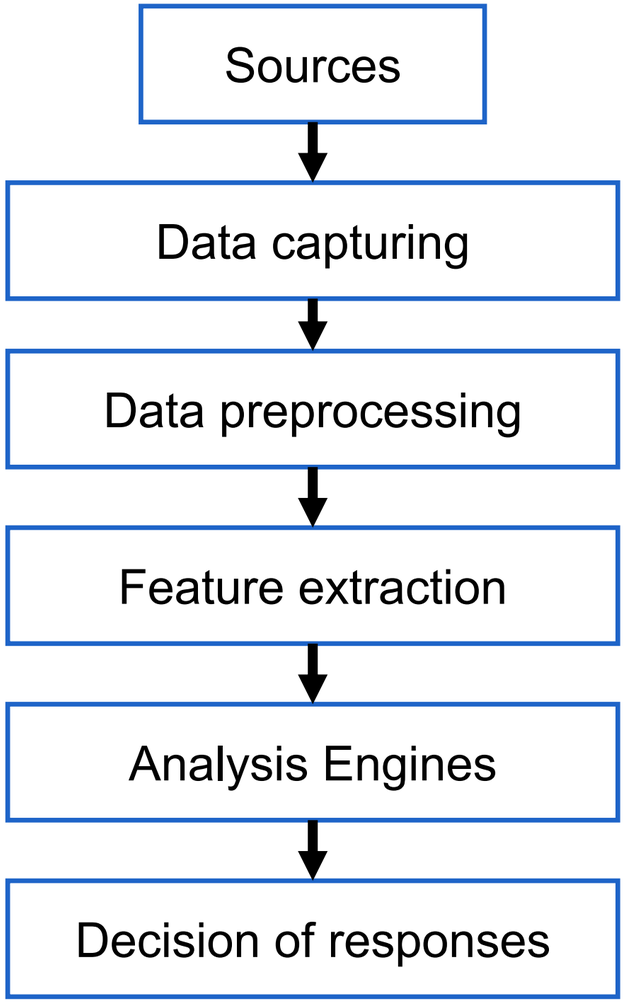
OVERVIEW


Sensors
Stored for offline processing
Data Stream
Pattern Rule Extraction
Complex Events
Derive Root Cause
Knowledge Base
Historical Data
Data Preprocessing
Profile Learning
Anomaly Detection
Action Mechanism

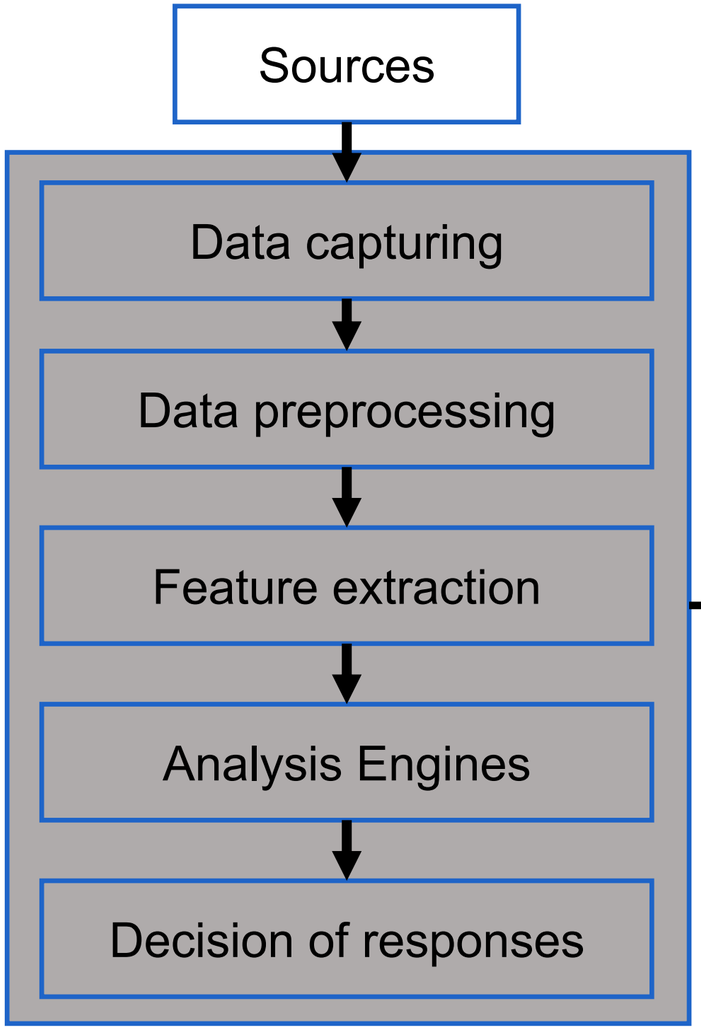
KNOWLEDGE INSIDE ML-BASED AD



ML PROCESS

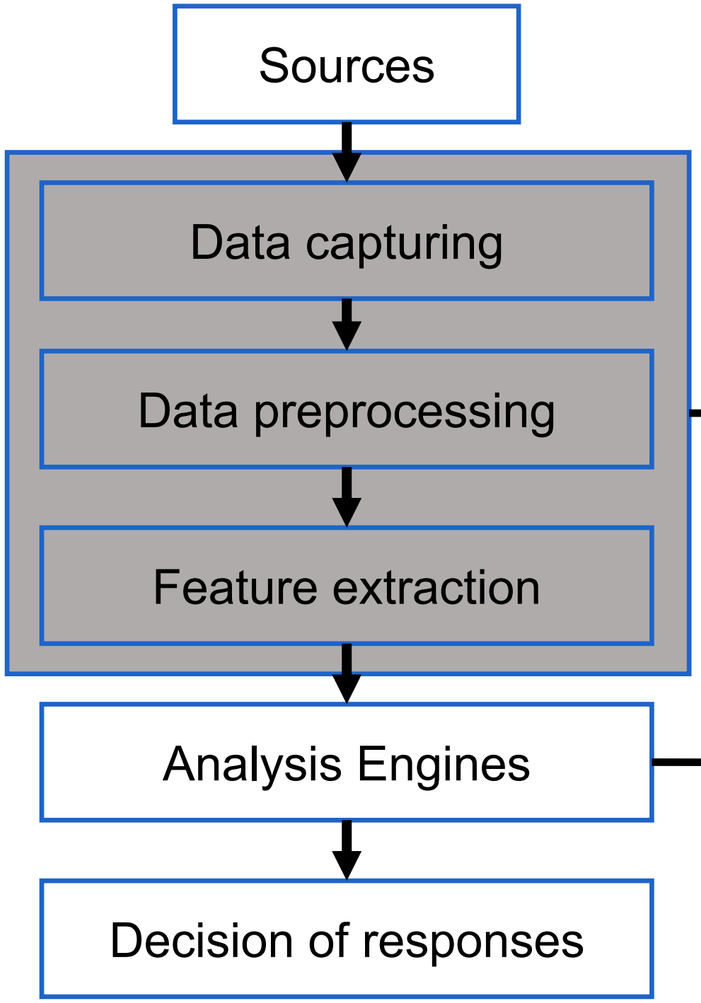

Sources
Data capturing
Data preprocessing
Feature extraction
Analysis Engines
Decision of responses

SOTA: KNOWLEDGE INSIDE ML-BASED AD



- Relational Graph Convolutional Network
- Inductive Logic Programming

SOTA: KNOWLEDGE INSIDE ML-BASED AD

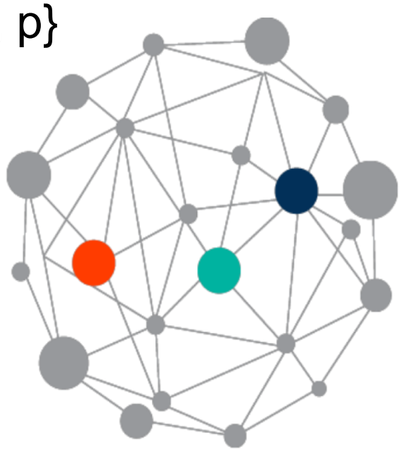


Domain Expertise:
- Knowledge graphs

Detection Method:
- Isolation forests
- Neural networks
- One-class SVM
- …

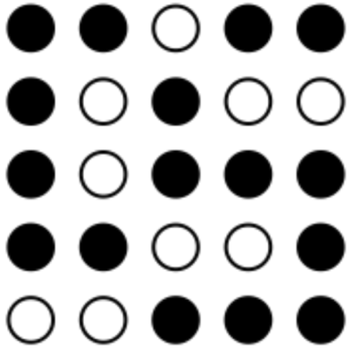
SOTA: EMBEDDING

{s, o, p}



Knowledge graph

Transformation



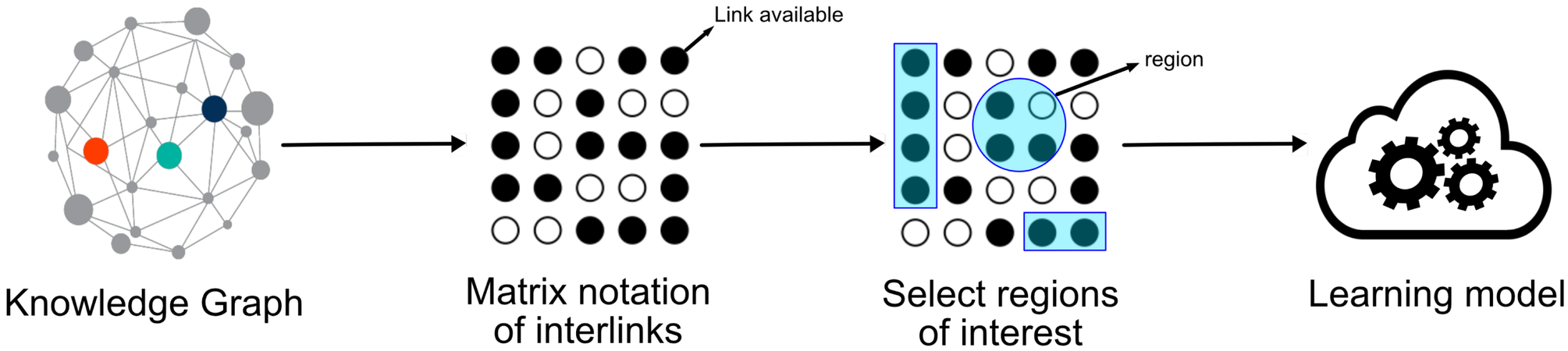
Matrix notation of interlinks

Transformation

0.2
-0.1
0.3
0.0
0.1
-0.3
-0.8
0.5

Vector

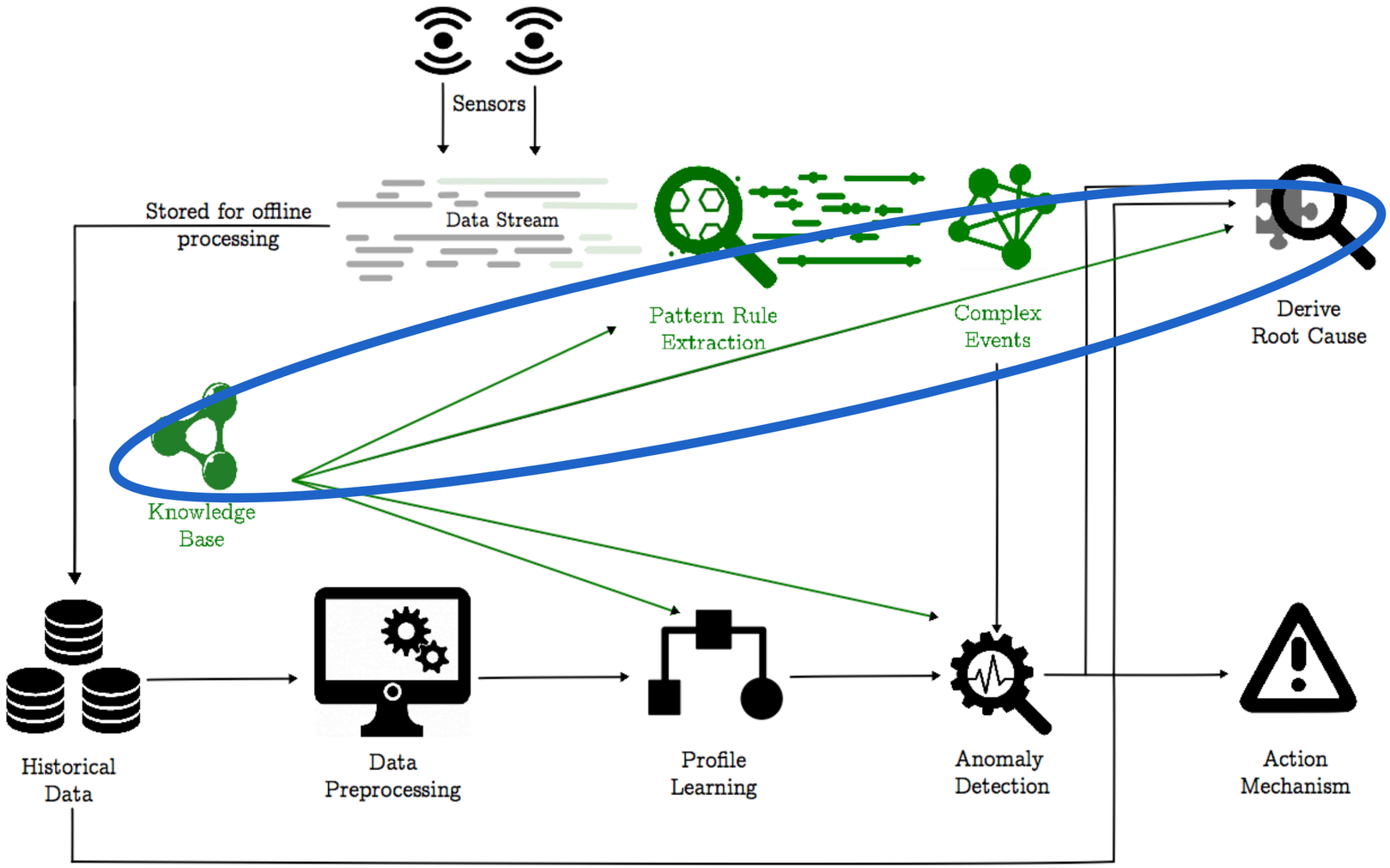
IMPROVED FEATURE GENERATION



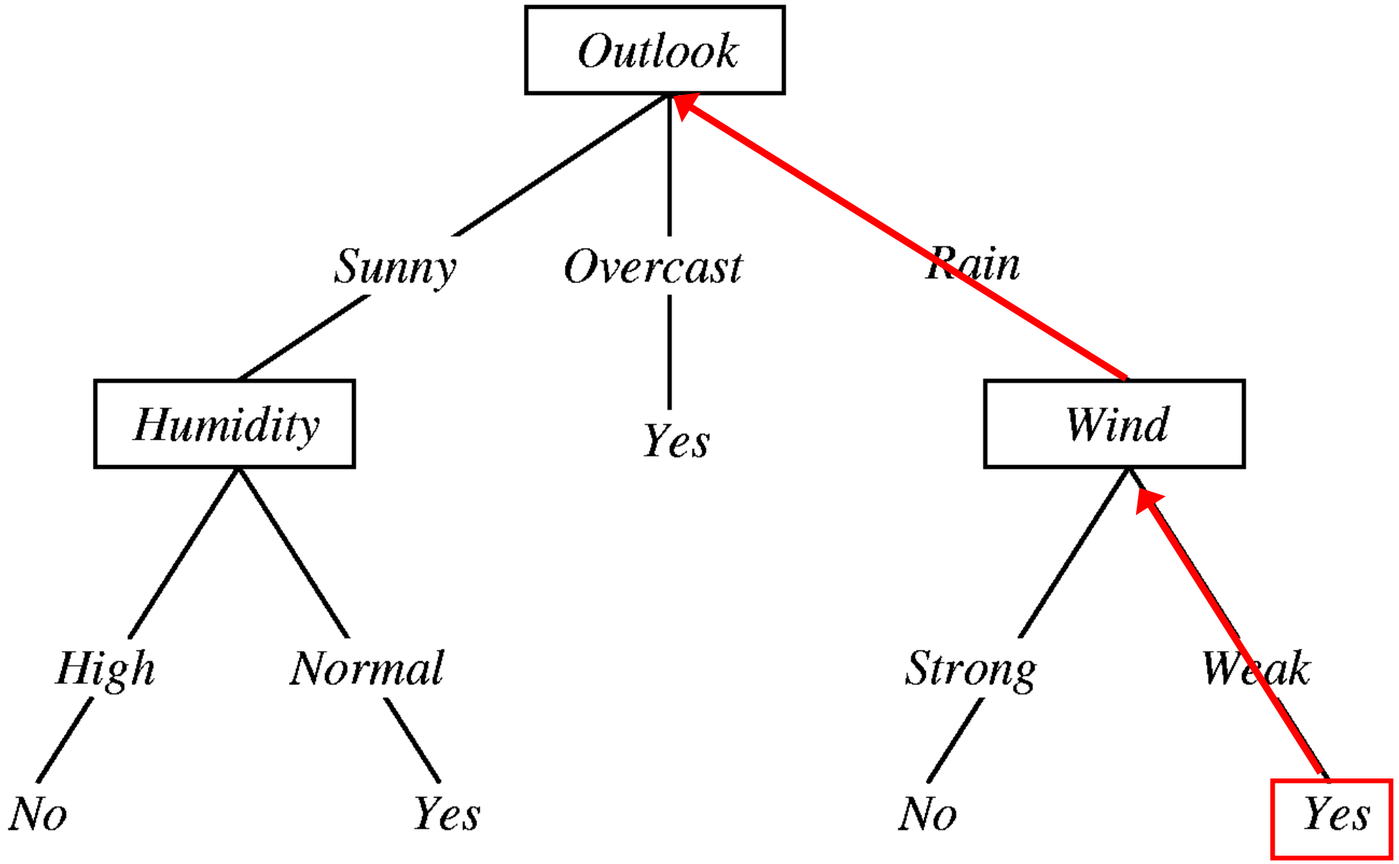
OUTLINE

1. Introduction
2. Feature importance with knowledge learning
3. Interpretable knowledge for cause detection
4. Adaptive stream detection
5. Additions & conclusion

CAUSE DETECTION



SOTA: CAUSE DETECTION



SOTA: CAUSE DETECTION

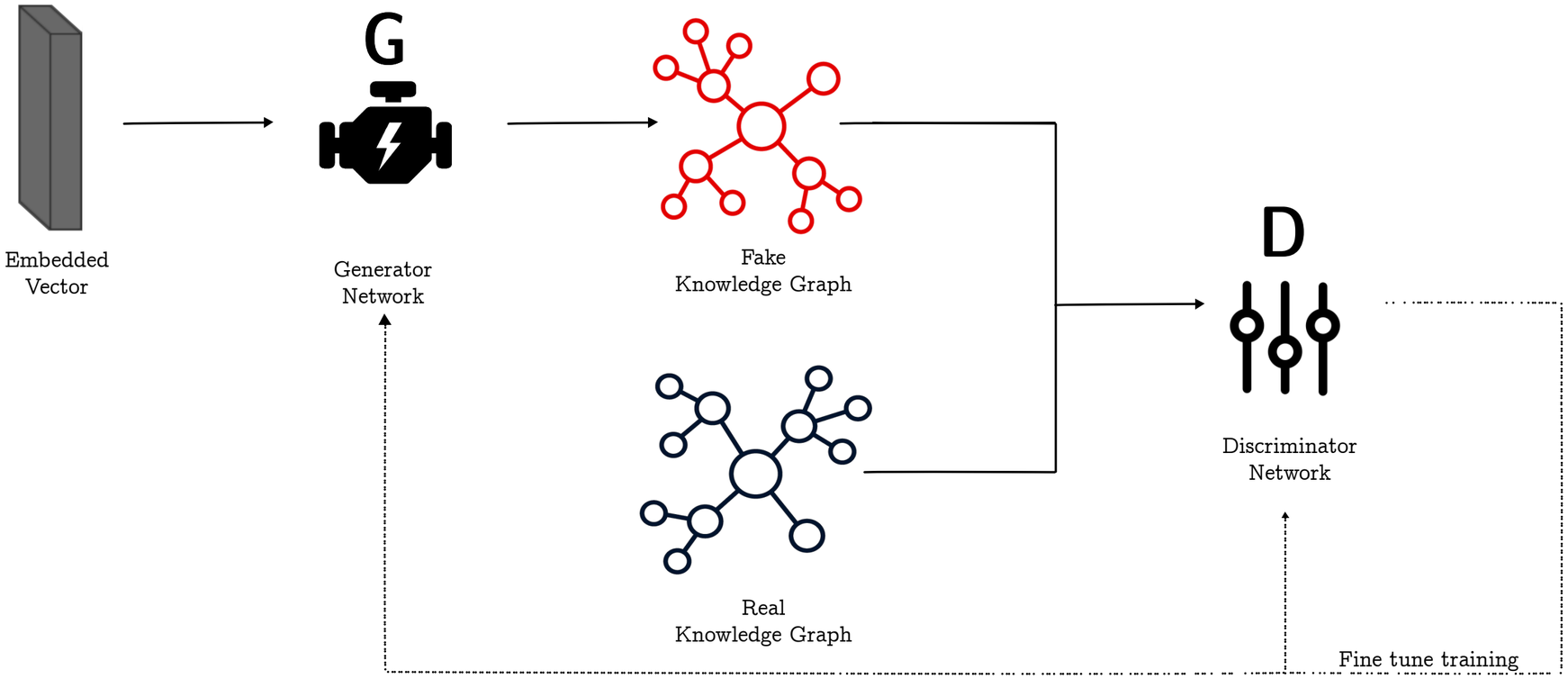
But:

- Models not always trees…
- Features not always interpretable (embeddings)
- Causes fixed (only from learned behavior)

=> Original knowledge graphs have however high interpretability and reasoning capacities

CAUSES FROM KNOWLEDGE

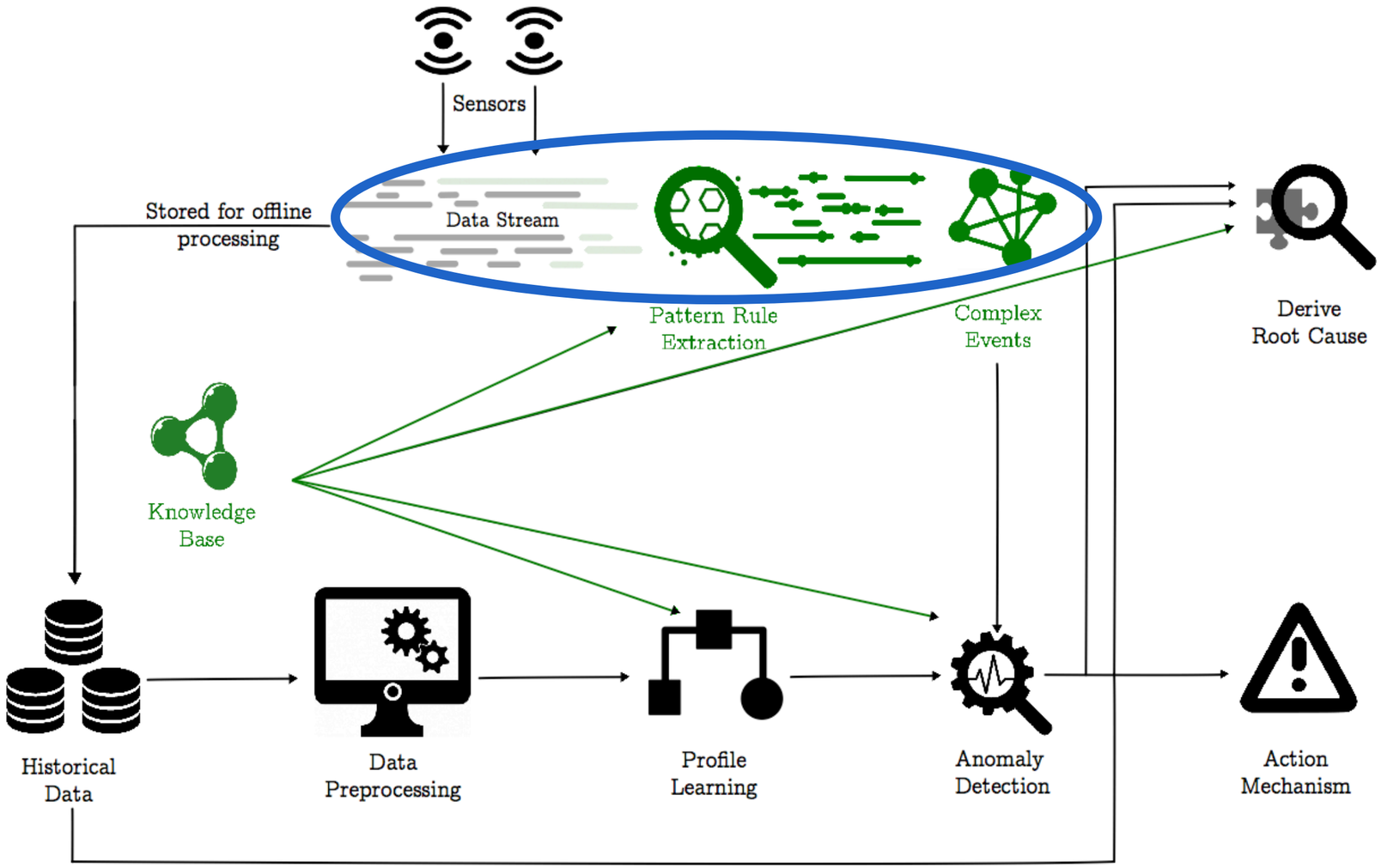
Generative Adversarial Network transforming the embeddings back to knowledge “graphs”



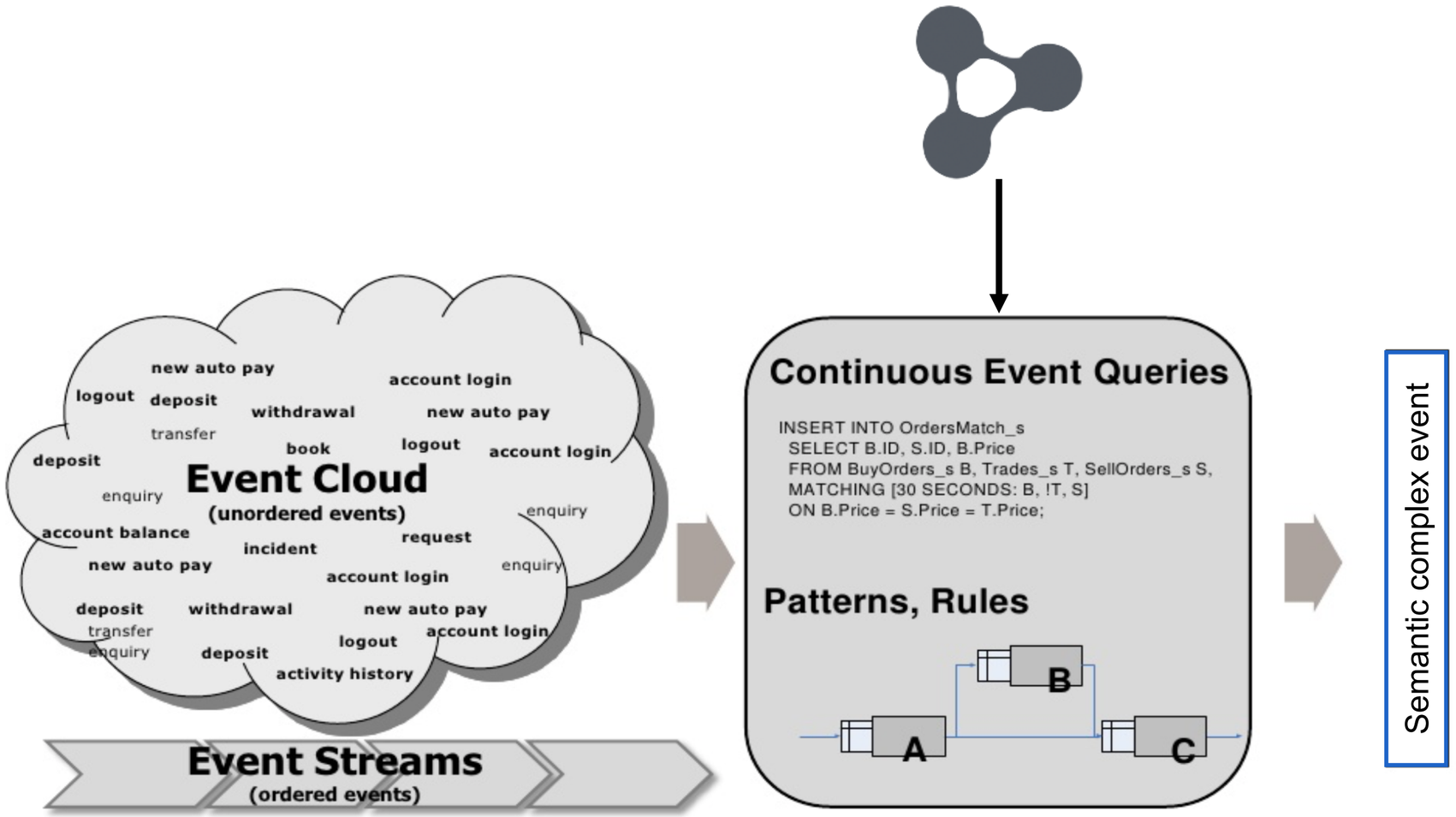
OUTLINE

1. Introduction
2. Feature importance with knowledge learning
3. Interpretable knowledge for cause detection
4. Adaptive stream detection
5. Additions & conclusion

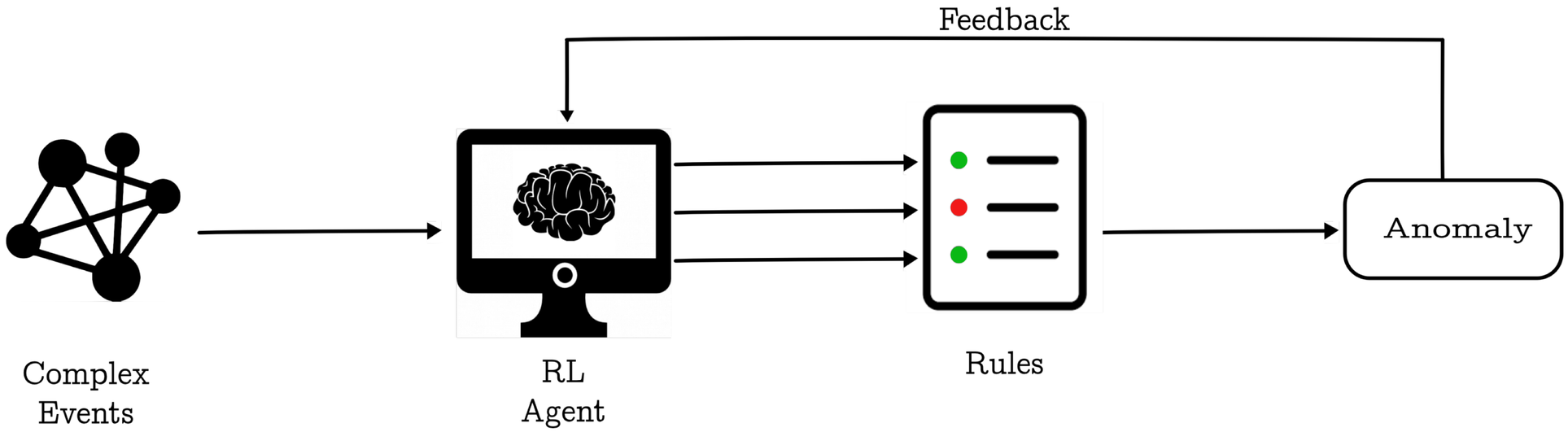
KNOWLEDGE INSIDE RULE-BASED SYSTEMS



SOTA: COMPLEX EVENT PROCESSING



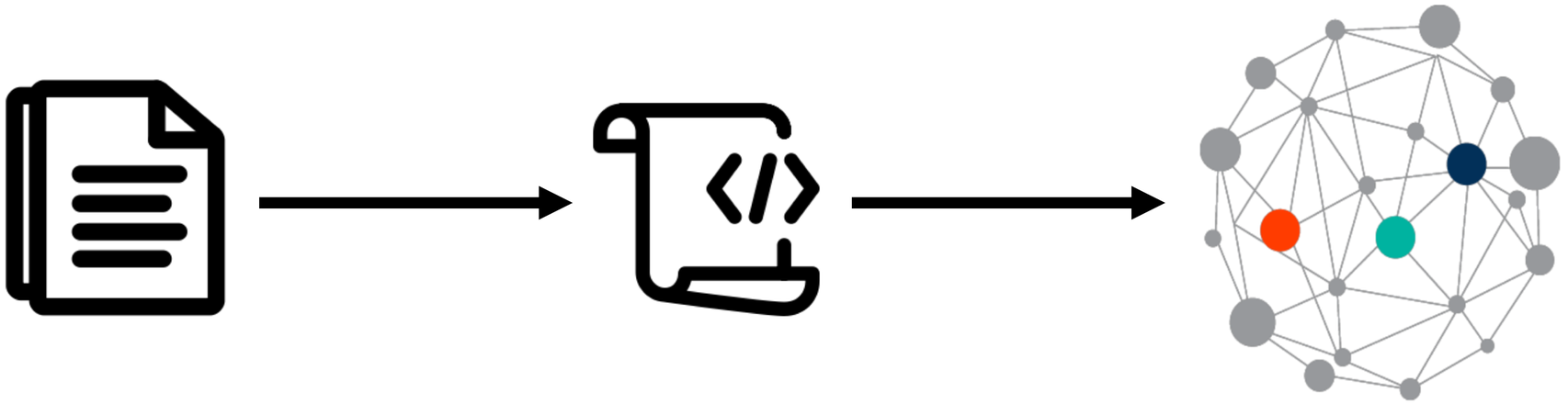
ADAPTIVE STREAM DETECTION



OUTLINE

1. Introduction
2. Feature importance with knowledge learning
3. Interpretable knowledge for cause detection
4. Adaptive stream detection
5. Additions & conclusion

FROM KNOWLEDGE TO GRAPH

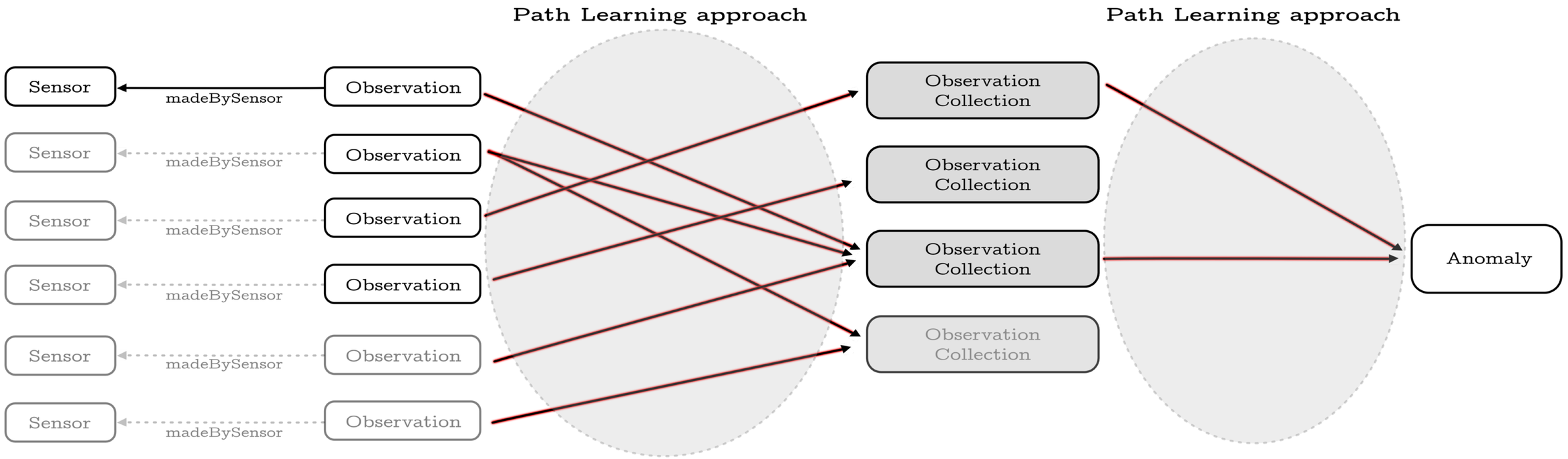


Informal documents

Script

Knowledge graph

LEARNING THE CAUSE OF AN OBSERVATION



CONCLUSION

- Hypothesis:
  1. Outperform detection rate by incorporating knowledge

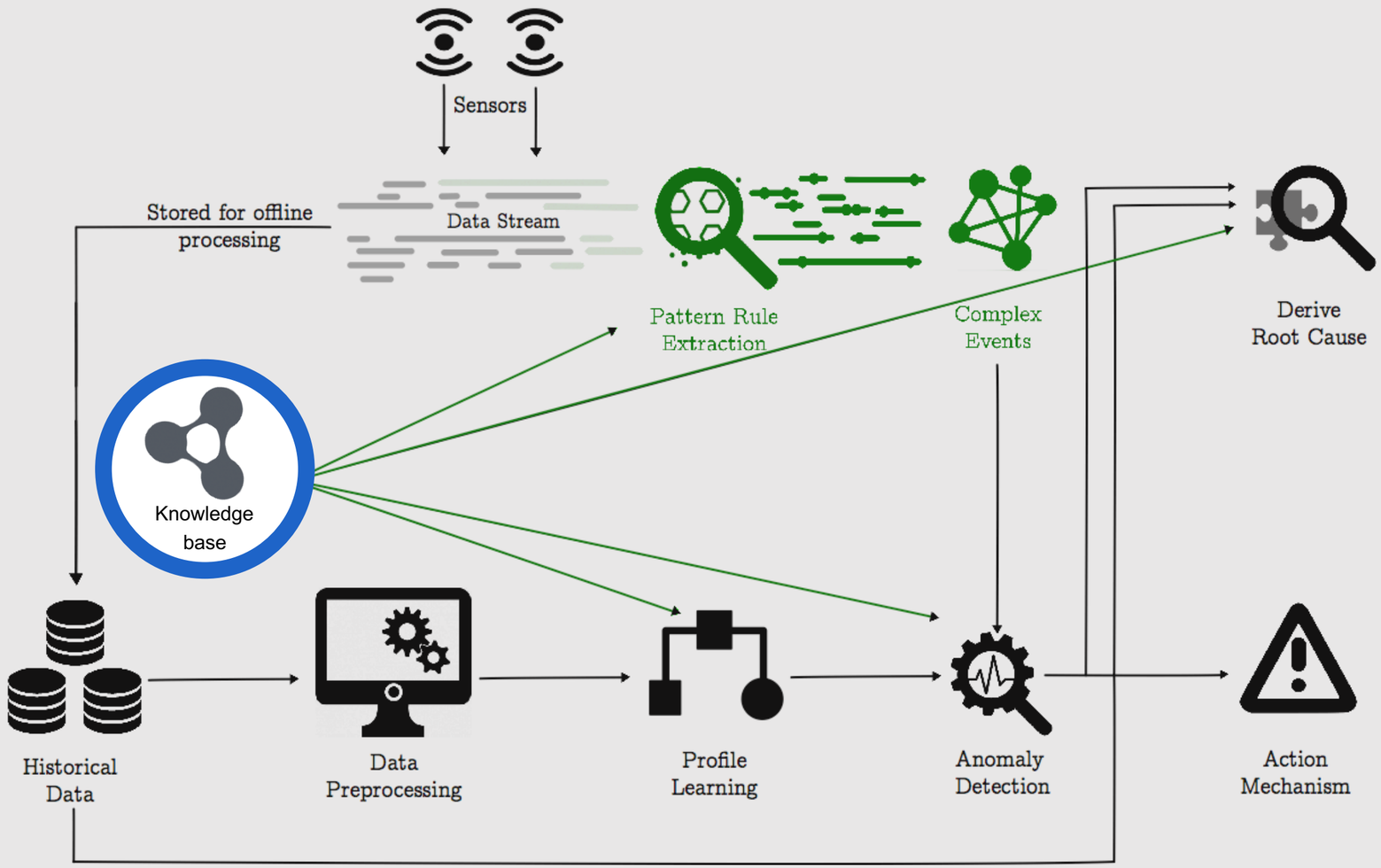
     => by extracting the knowledge in a closed loop with the models
  2. Reduce F1 score, the number of false detects, with knowledge

     => by limiting the effect of the frequently used accuracy metric
  3. Reduce the human involvement drastically

     => By deriving the causes from the available knowledge
  4. Make techniques available for streaming environments

     => By adaptively adding the model-based rules

CONCLUSION


Sensors
Stored for offline processing
Data Stream
Pattern Rule Extraction
Complex Events
Derive Root Cause
Knowledge base
Historical Data
Data Preprocessing
Profile Learning
Anomaly Detection
Action Mechanism

THANK YOU!

Acknowledgements:
- Reviewers & organizing committee
- My mentor: Anna Lisa Gentile
- My promotors: Filip De Turck & Femke Ongenae
- Industry partner: imec, Televic & Renson


bram.steenwinckel@ugent.be

@bsteenwin