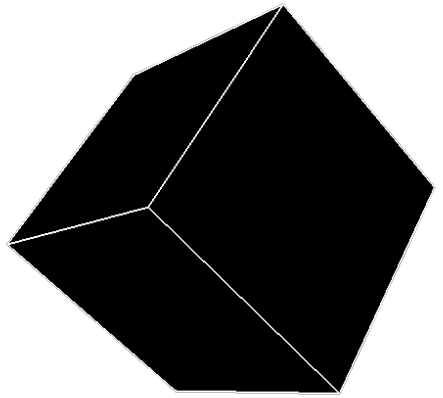
Explainable Machine Learning: Mind the Users and their Knowledge

Freddy Lécué
Inria, France
CortAIx@Thales, Canada
@freddylecue

Intelligent Systems Conference (IntelliSys) 2020

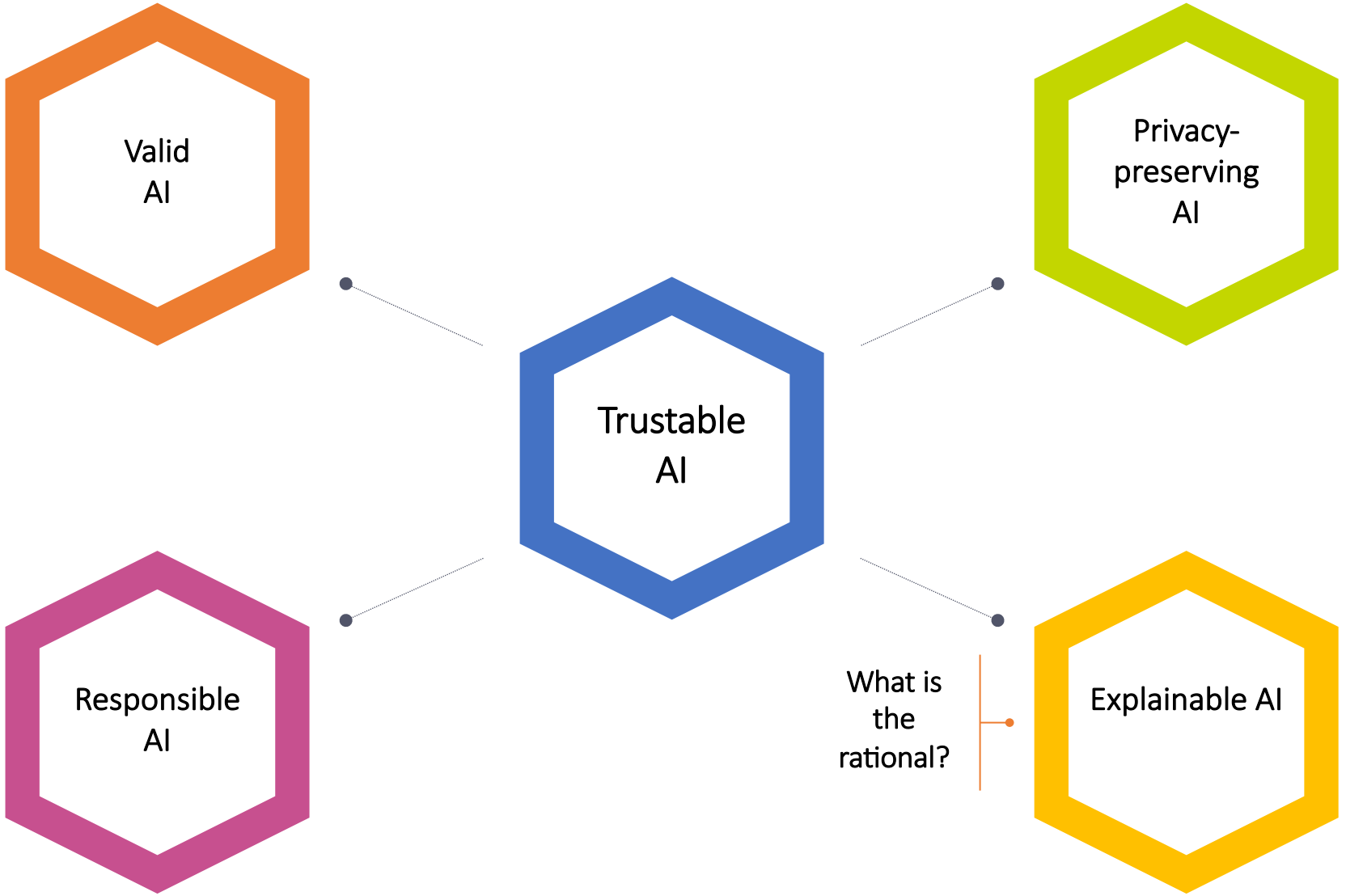
September 4th, 2020
Amsterdam, The Netherlands



Freddy Lecue. On the Role of Knowledge Graphs in Explainable AI. Semantic Web Journal (2020) http://www.semantic-web-journal.net/content/role-knowledge-graphs-explainable-ai

Scope

AI Adoption: Requirements



Explanation in AI

Explanation in AI aims to create a suite of techniques that produce more explainable models, while maintaining a high level of searching, learning, planning, reasoning performance: optimization, accuracy, precision; and enable human users to understand, appropriately trust, and effectively manage the emerging generation of AI systems .

Motivation

Business to Customer

Gary Chavez added a photo you might be in. ...

about a minute ago ·

Critical Systems


BNSF

Markets We Serve (Critical Systems)

Aerospace

Space

Ground Transportation

Defence

Security

Trusted Partner For A Safer World

But not Only Critical Systems

COMPAS recidivism black bias

Opinion

OP-ED CONTRIBUTOR

When a Computer Program Keeps You in Jail

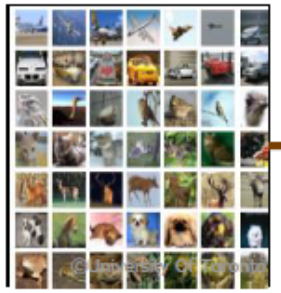
By Rebecca Wexler

June 13, 2017

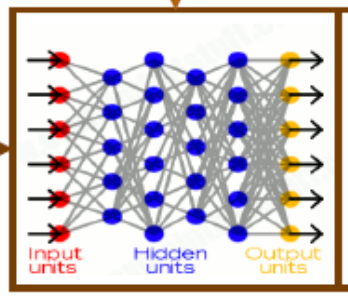
XAI in a Nutshell

XAI in a Nutshell

Today



Learning Process

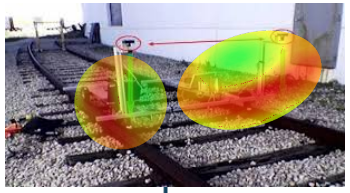


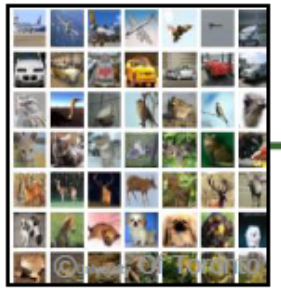
This is an obstacle on rail train

- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

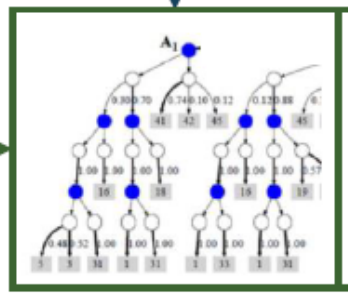
Training Data | Learned Function | Output | User with a Task

Tomorrow





New Learning Process



Obstacle on rail train

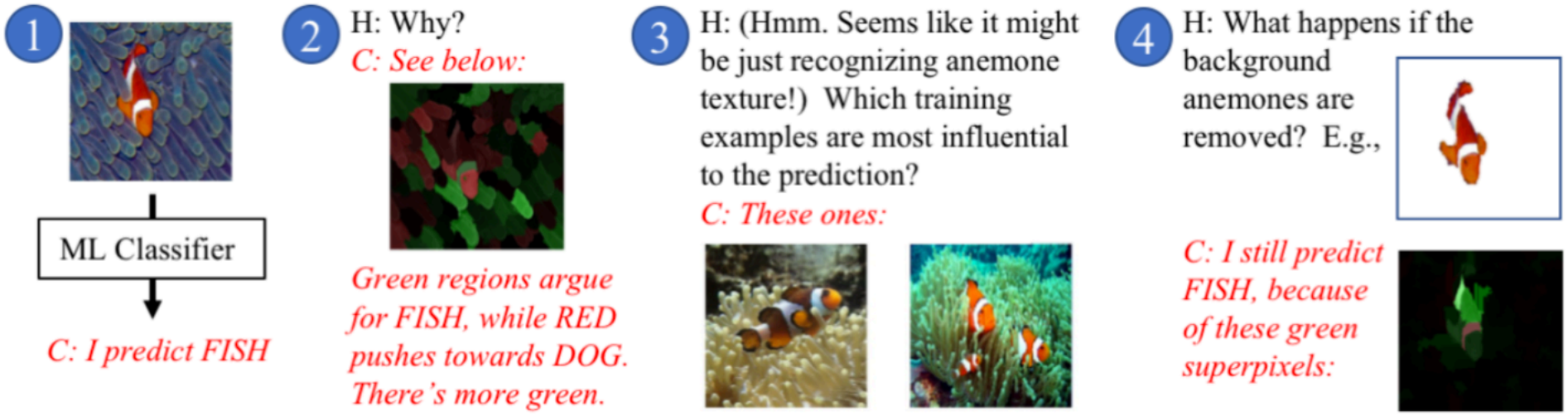
- **Obstruction covering full width**

- I understand why
- I understand why not
- I know when you'll succeed
- I know when you'll fail
- I know when to trust you
- I know why you erred

Training Data | Explainable Model | Explanation Interface | User with a Task

Source: https://www.cc.gatech.edu/~alanwags/DLAI2016/(Gunning)%20IJCAI-16%20DLAI%20WS.pdf

An Example of an end-to-end XAI System



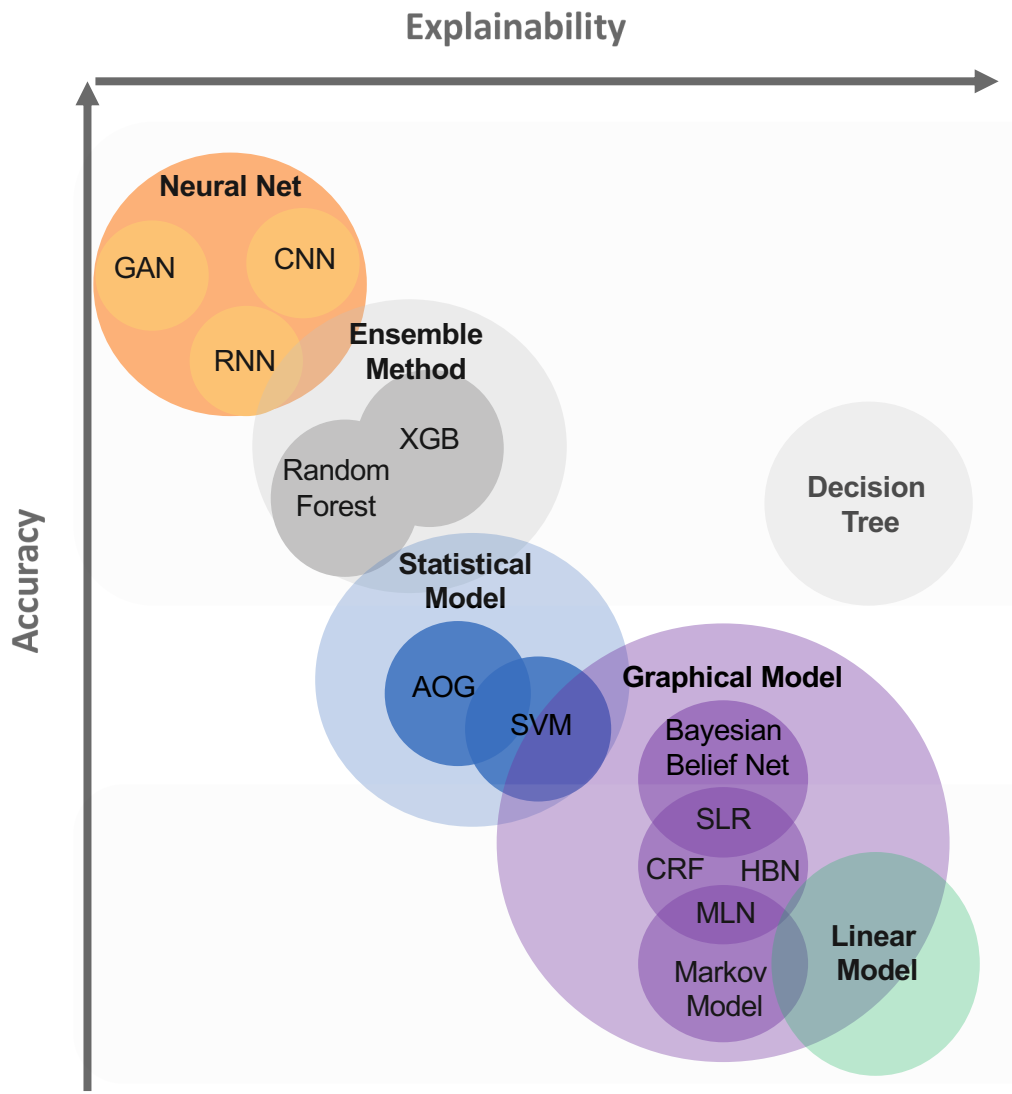
- Humans may have follow-up questions
- Explanations cannot answer all users' concerns

Weld, D., and Gagan Bansal. "The challenge of crafting intelligible intelligence." Communications of ACM (2018).

How to Explain? Accuracy vs. Explanability

Learning

- Challenges:
  - Supervised
  - Unsupervised learning
- Approach:
  - Representation Learning
  - Stochastic selection
- Output:
  - **Correlation**
  - **No causation**



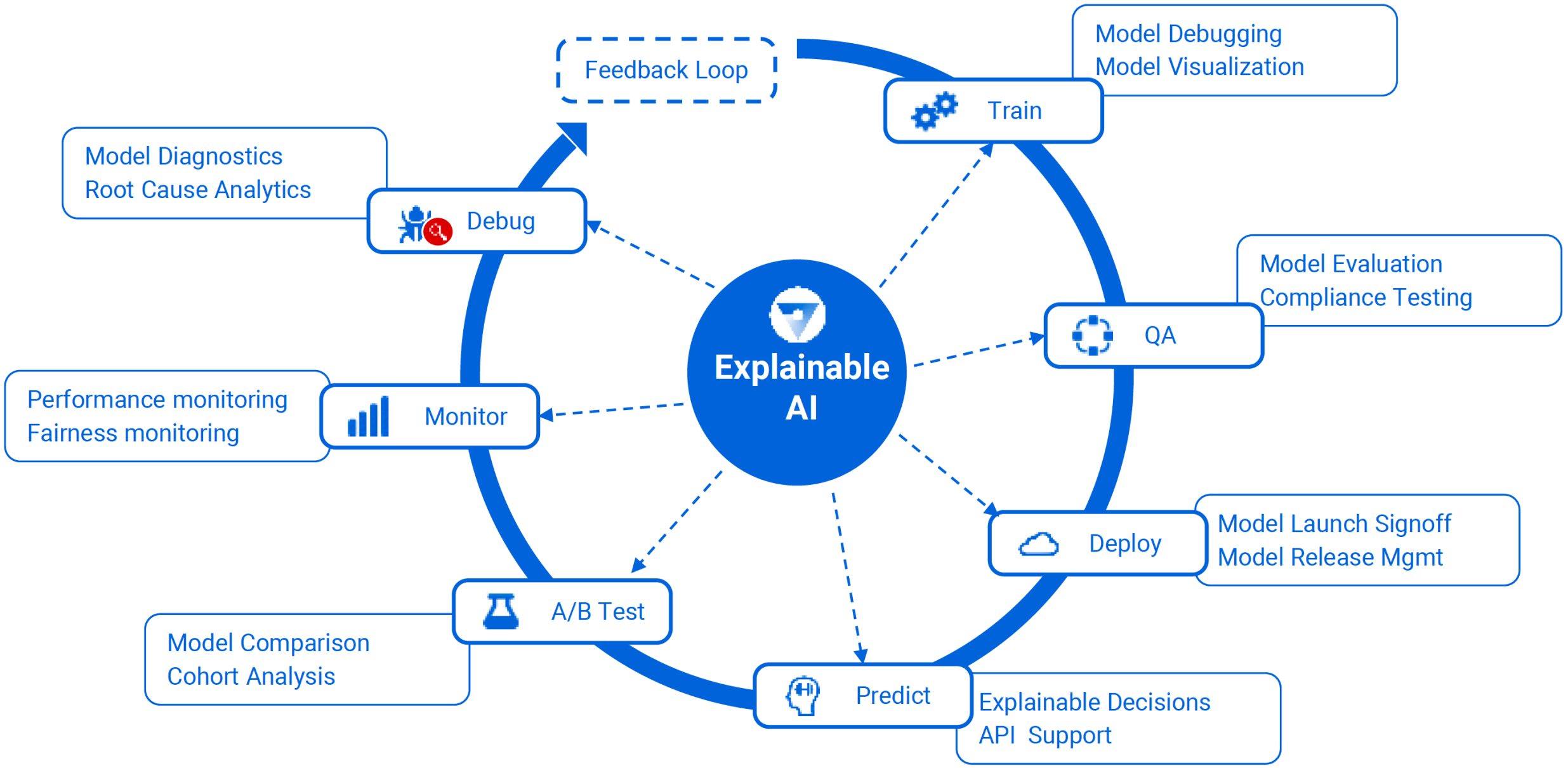
Interpretability

- Non-Linear functions
- Polynomial functions
- Quasi-Linear functions

XAI Objective

Supporting Industrialization of AI at Scale

Explainability by Design for AI Products



On Role of Data In XAI

Interpretable Data for Interpretable Models

Table of baby-name data
(baby-2010.csv)

name	rank	gender	year
Jacob	1	boy	2010
Isabella	1	girl	2010
Ethan	2	boy	2010
Sophia	2	girl	2010
Michael	3	boy	2010

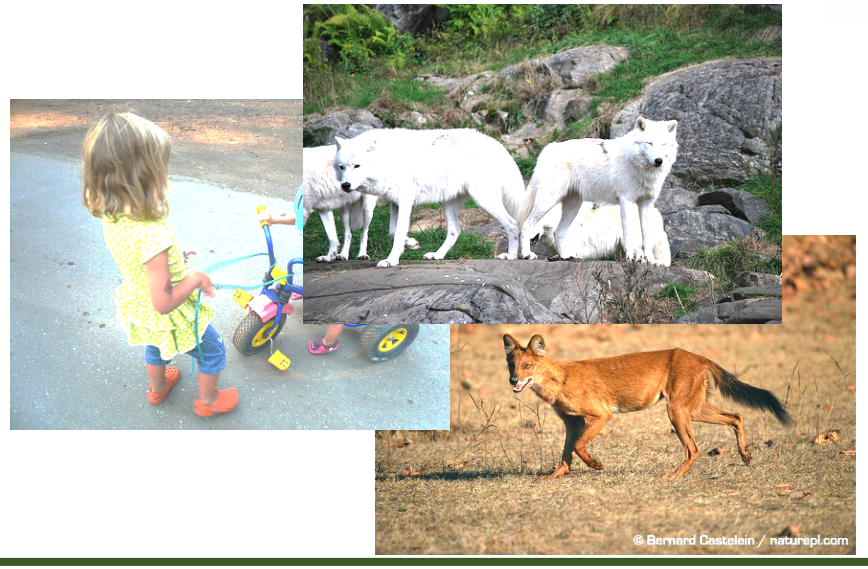
Field names

One row (4 fields)

2000 rows all told

Tabular

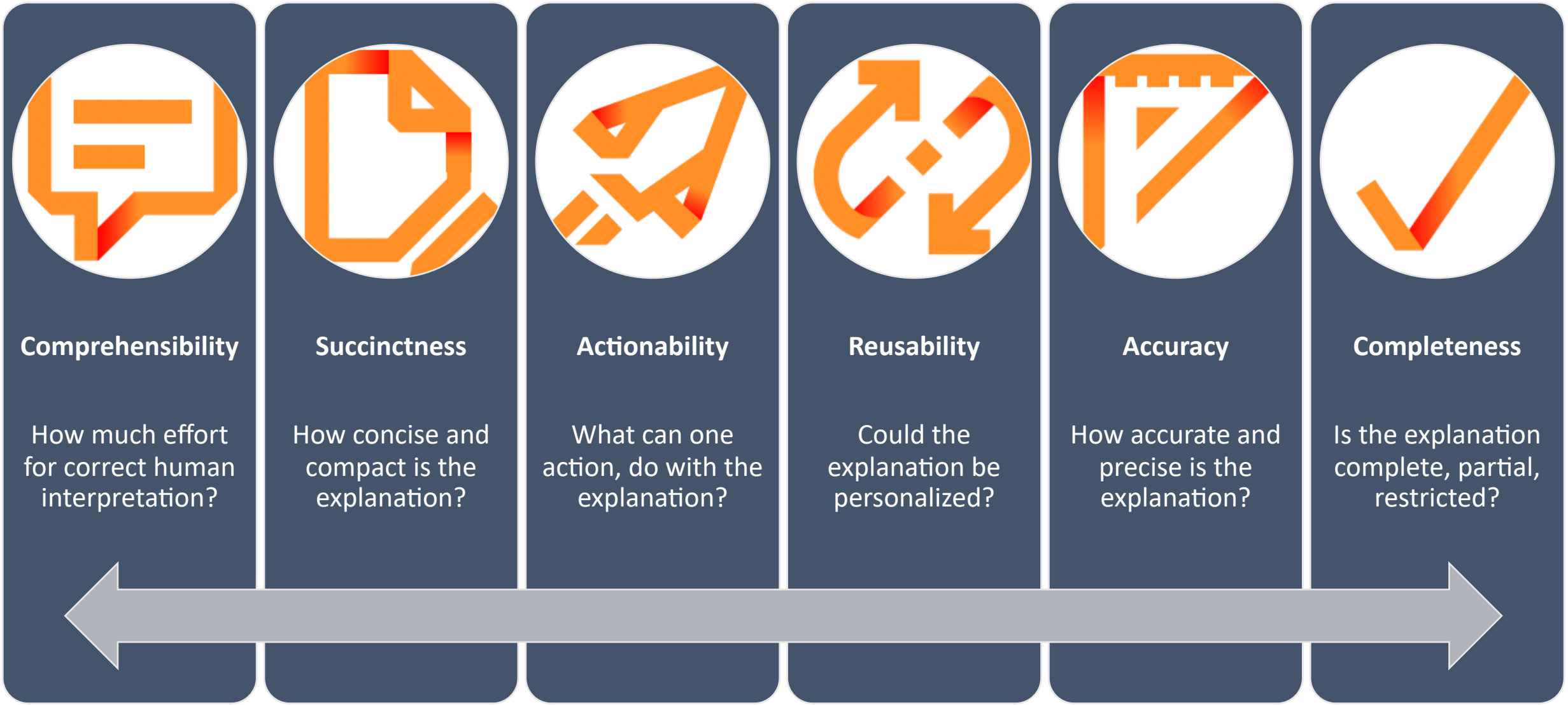
Images



Text

What about the Evaluation?

XAI: One Objective, Many Metrics



Source: Accenture Point of View. Understanding Machines: Explainable AI. Freddy Lecue, Dadong Wan

XAI in AI

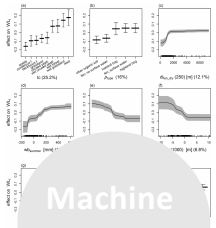
XAI: One Objective, Many 'AI's, Many Definitions, Many Approaches

How to summarize the reasons (motivation, justification, understanding) for an AI system behavior, and explain the causes of their decisions?

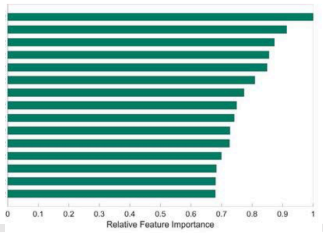
Artificial Intelligence

Machine Learning

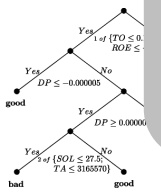
Dependency Plot



Feature Importance



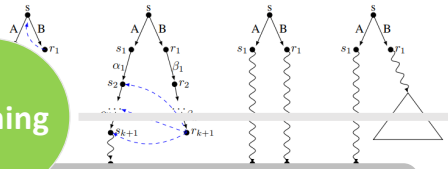
Surrogate Model



Which features are responsible of classification?

Planning

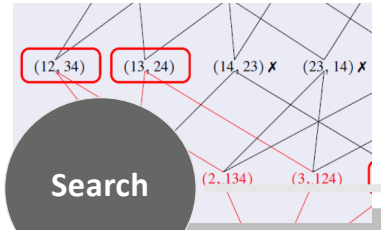
Plan Refinement



Which actions are responsible of a plan?

Search

Conflicts Resolution



Which constraints can be relaxed?

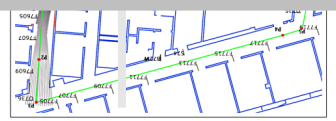
Game Theory

Which combination of features is optimal?

Shapely Values

Robotics

Which decisions, combination of multimodal decisions lead to an action?



Narrative-based

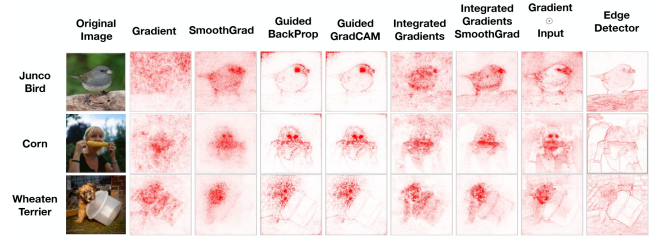
MAS

Strategy Summarization

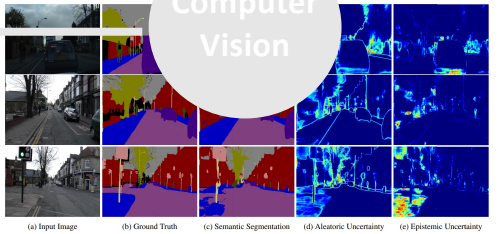
- Which agent strategy & plan ?
- Which player contributes most?
- Why such a conversational flow?

Computer Vision

Saliency Map



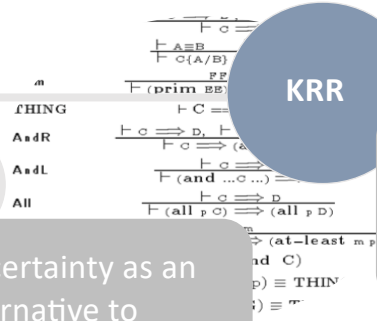
Which complex features are responsible of classification?



Uncertainty Map

KRR

Diagnosis



Abduction

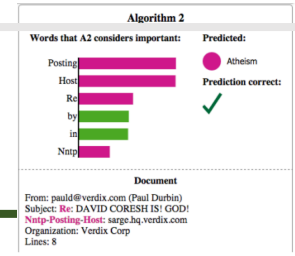
- Which axiom is responsible of inference (e.g., classification)?
- Abduction/Diagnostic: Find the **right** root causes (abduction)?

UAI

Uncertainty as an alternative to explanation

Machine Learning based

NLP



Which entity is responsible for classification?

On the Role of Knowledge Graphs in Explainable AI

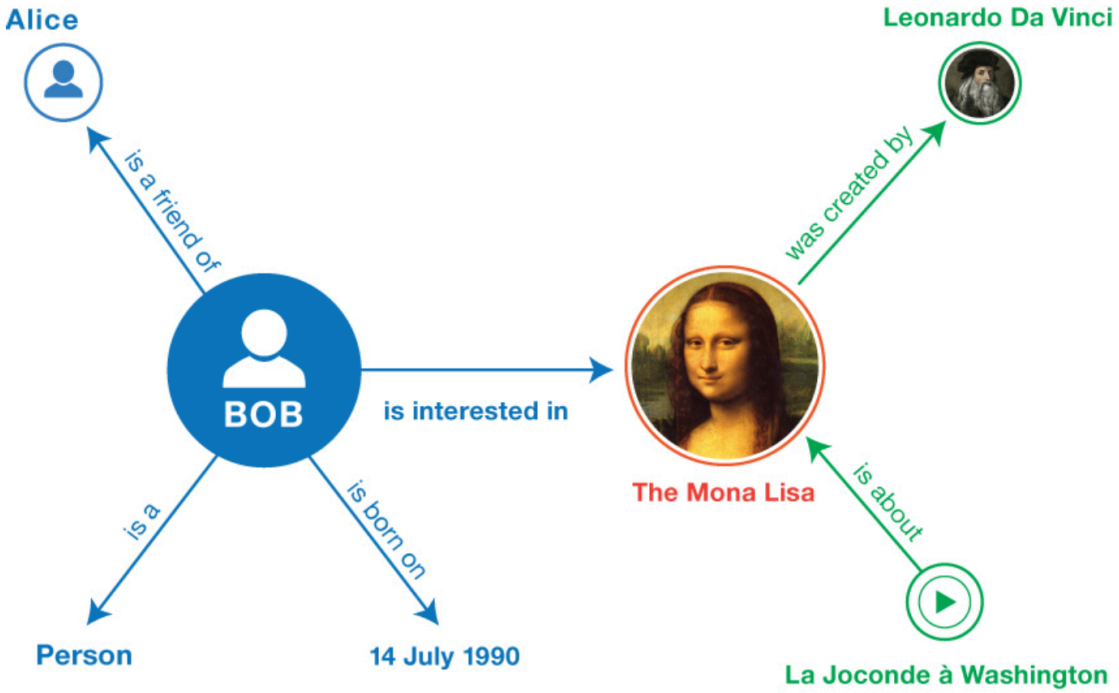
A Machine Learning Perspective

On the Role of Knowledge Graph in Explainable AI - Semantic Web Journal - http://www.semantic-web-journal.net/content/role-knowledge-graphs-explainable-ai

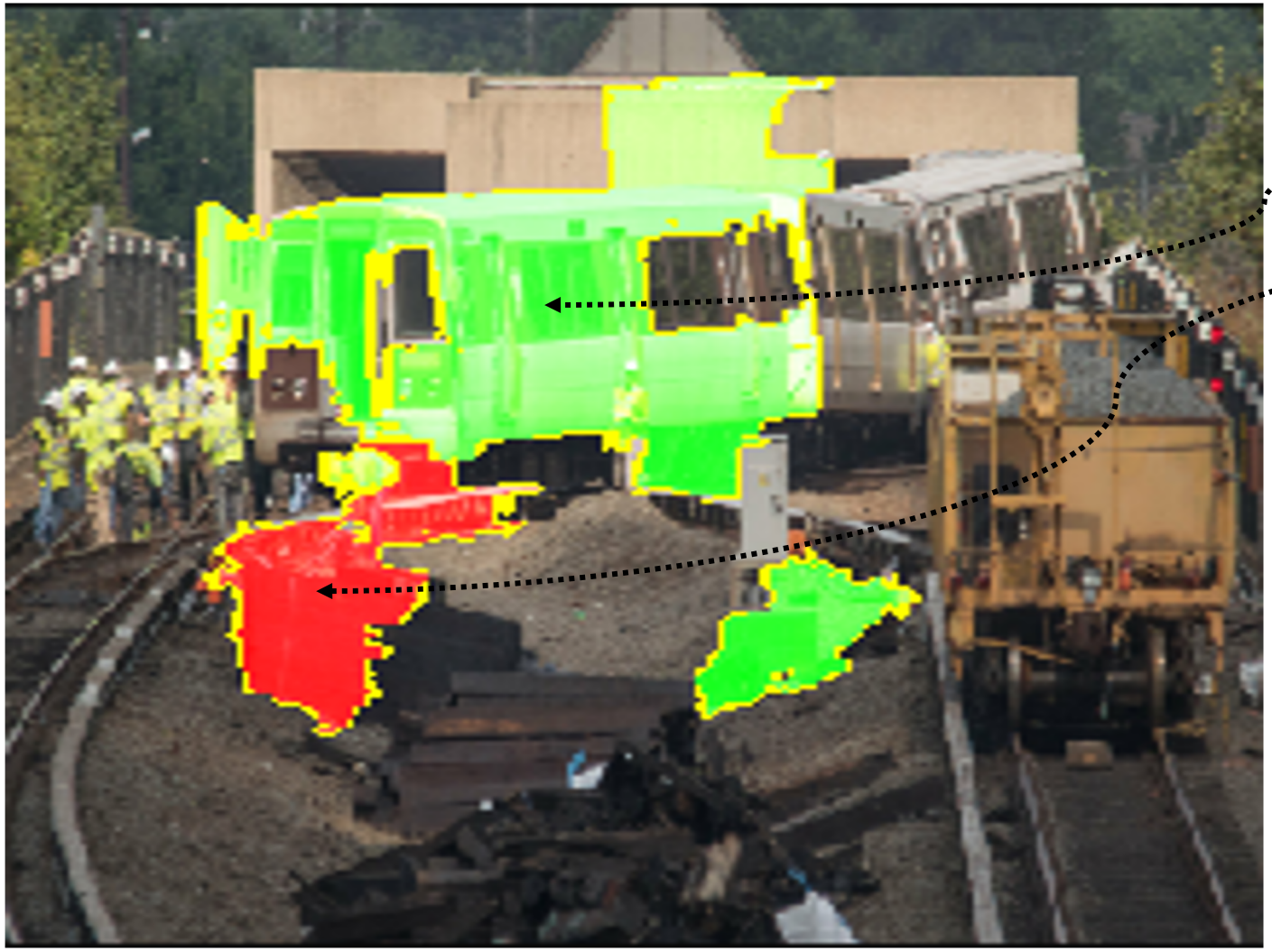
Knowledge Graph (1)

- Set of (*subject*, *predicate*, *object* — *SPO*) **triples** - *subject* and *object* are **entities**, and *predicate* is the **relationship** holding between them.
- Each SPO **triple** denotes a **fact**, i.e. the existence of an actual relationship between two entities.

subject	predicate	object
Bob	*is interested in*	*The Mona Lisa*
Bob	*is a friend of*	*Alice*
The Mona Lisa	*was created by*	*Leonardo Da Vinci*
Bob	*is a*	*Person*
La Joconde à W.	*is about*	*The Mona Lisa*
Bob	*is born on*	*14 July 1990*



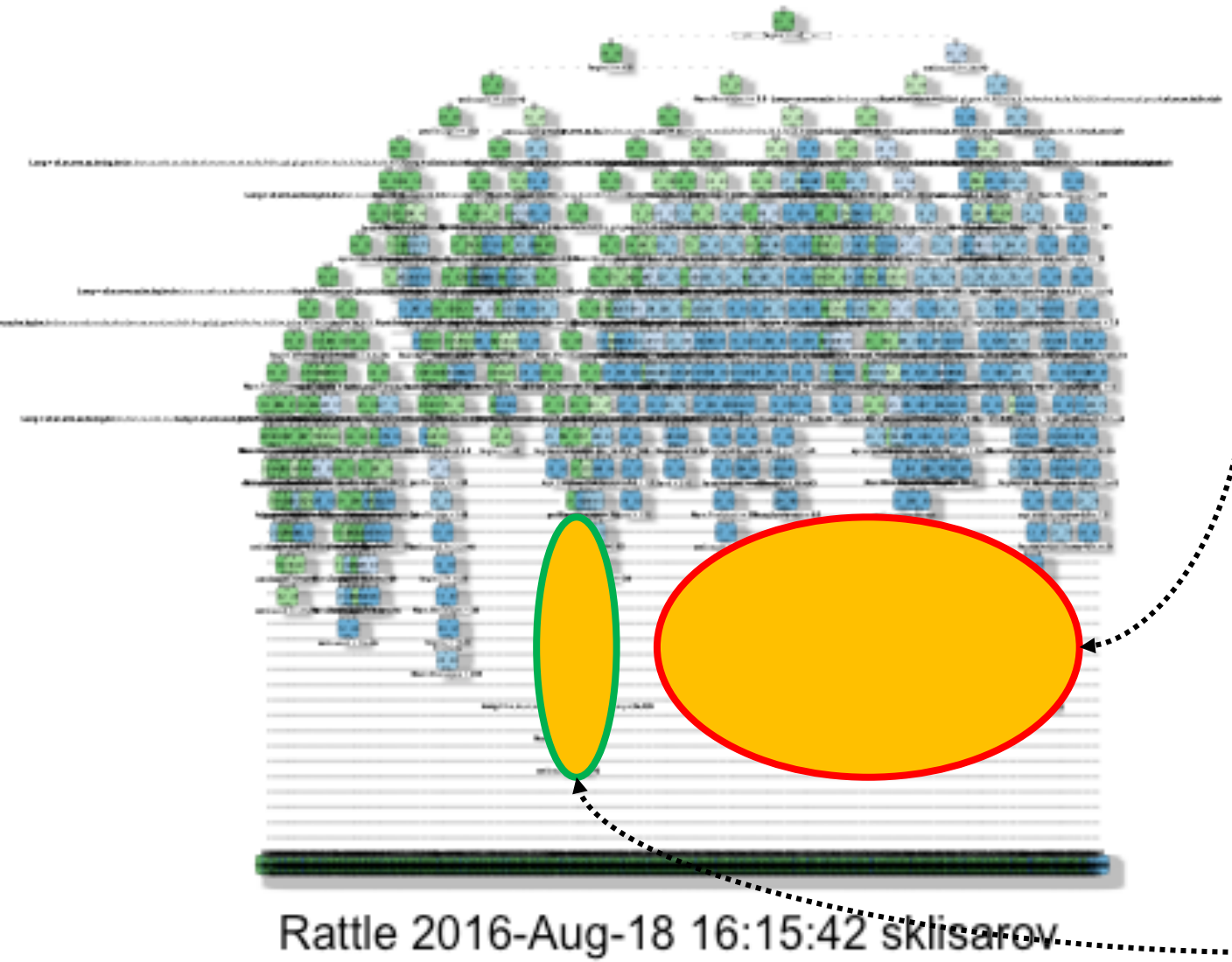
Knowledge Graph in Machine Learning (1)



Augmenting (input) features with more semantics such as knowledge graph embeddings / entities

https://stats.stackexchange.com/questions/230581/decision-tree-too-large-to-interpret

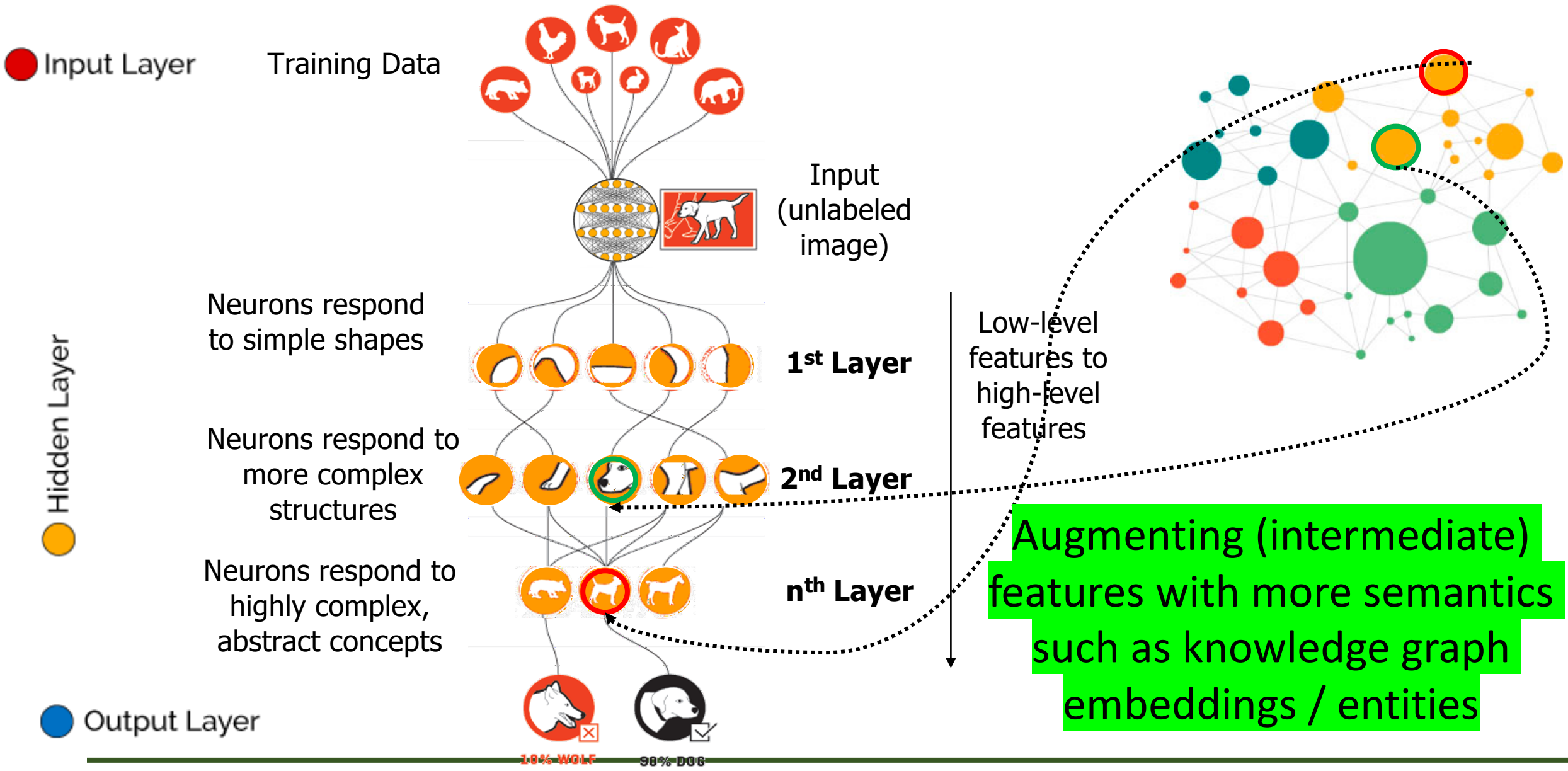
Knowledge Graph in Machine Learning (2)



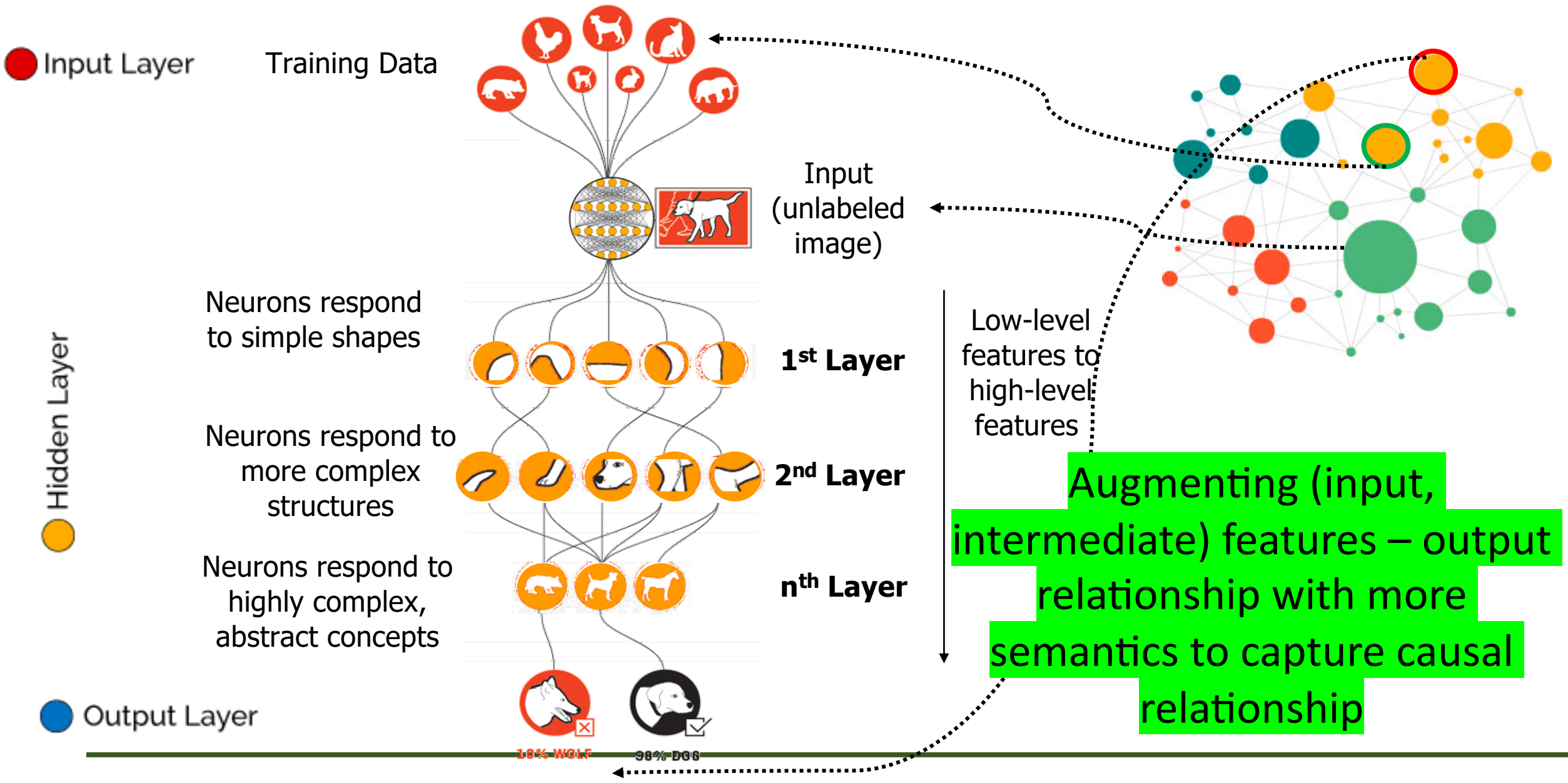
Augmenting machine learning models with more semantics such as knowledge graphs entities

https://stats.stackexchange.com/questions/230581/decision-tree-too-large-to-interpret

Knowledge Graph in Machine Learning (3)



Knowledge Graph in Machine Learning (4)



Knowledge Graph in Machine Learning (5)

Description 1: This is an orange train accident

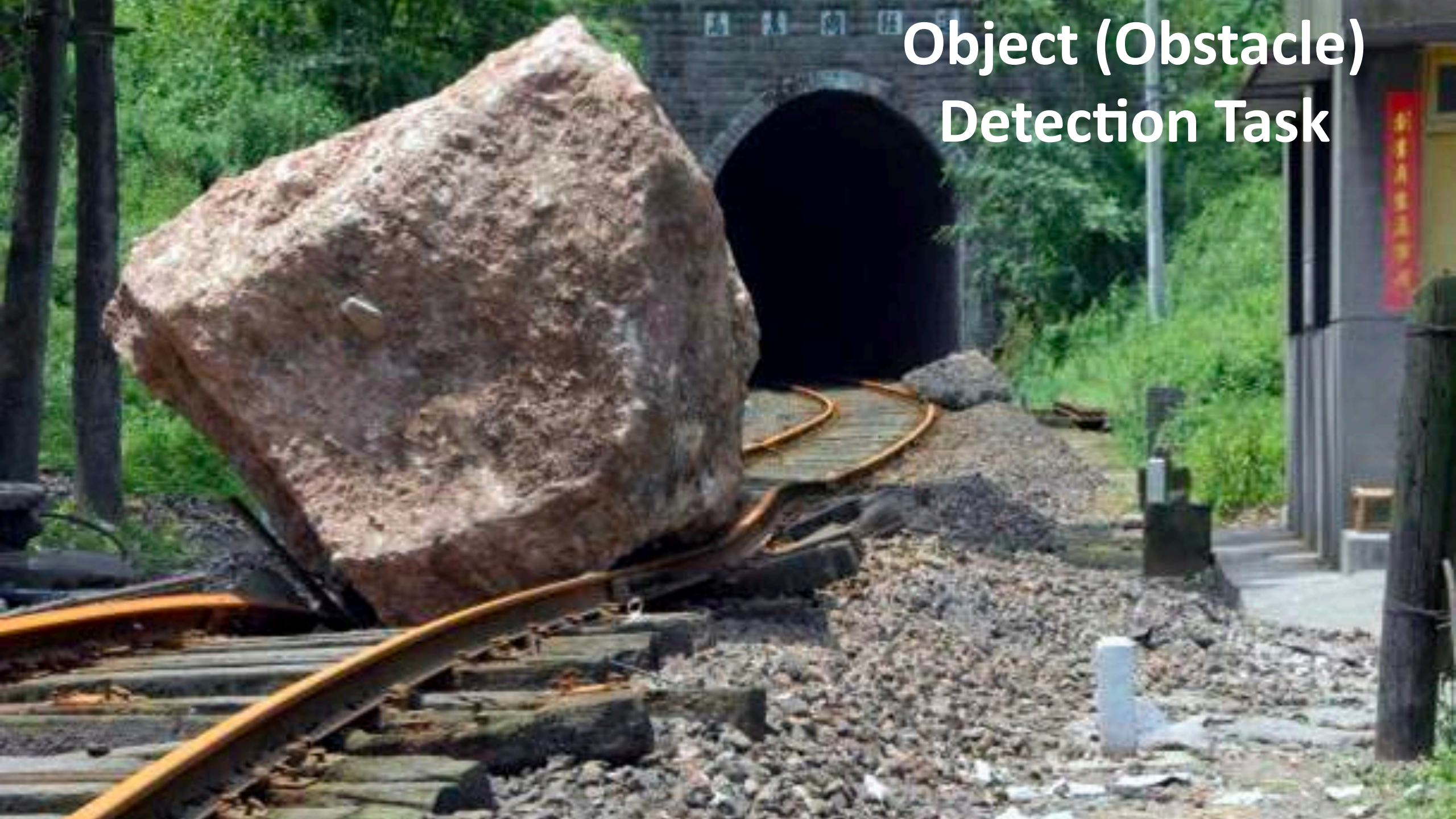
Description 2: This is an train accident between two speed merchant trains of characteristics X43-B and Y33-C in a dry environment

Description 3: This is a public transportation accident

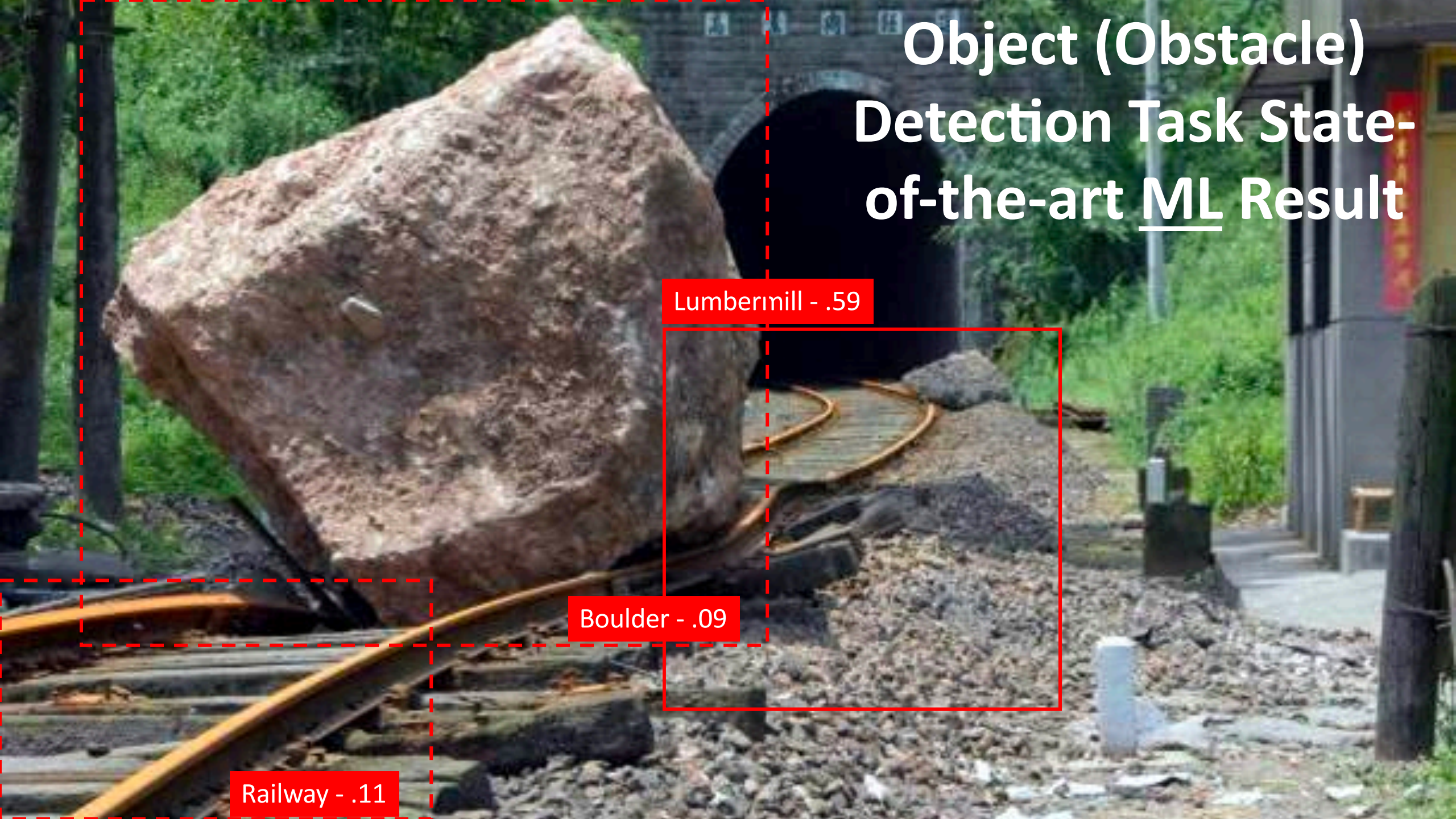
Augmenting models with semantics to support personalized explanation

On One Industrial Application in Thales

State of the Art Machine Learning Applied to Critical Systems

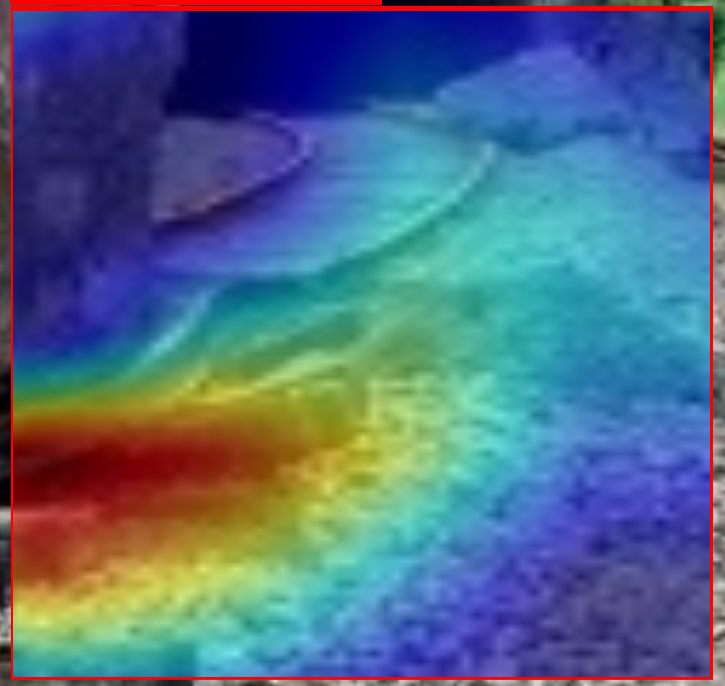

Object (Obstacle)
Detection Task

Object (Obstacle) Detection Task State-of-the-art ML Result


Object (Obstacle) Detection Task State-of-the-art ML Result
Lumbermill - .59
Boulder - .09
Railway - .11

State of the Art XAI Applied to Critical Systems

Object (Obstacle) Detection Task State-of-the-art XAI Result



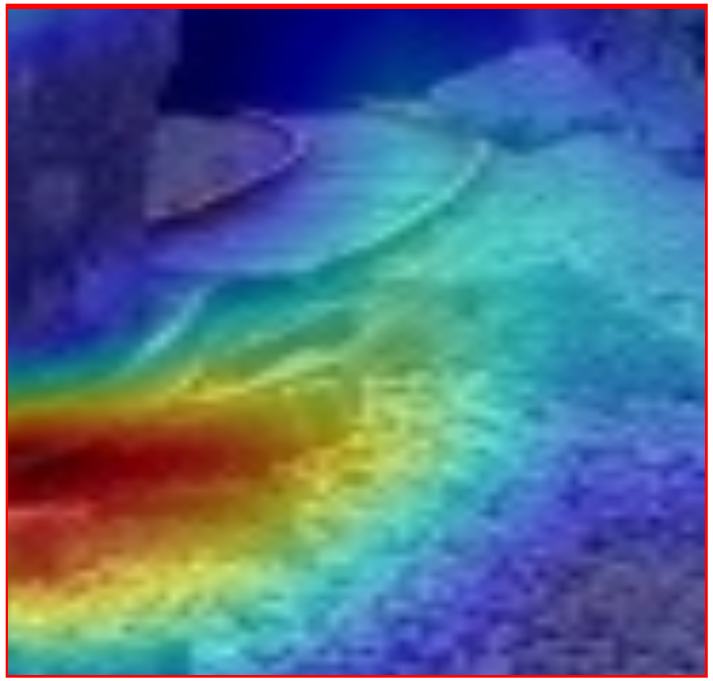
Unfortunately, this is of NO use for a human behind the system

Let’s stay back

Why this Explanation?
(meta explanation)

After Human Reasoning…

Lumbermill - .59

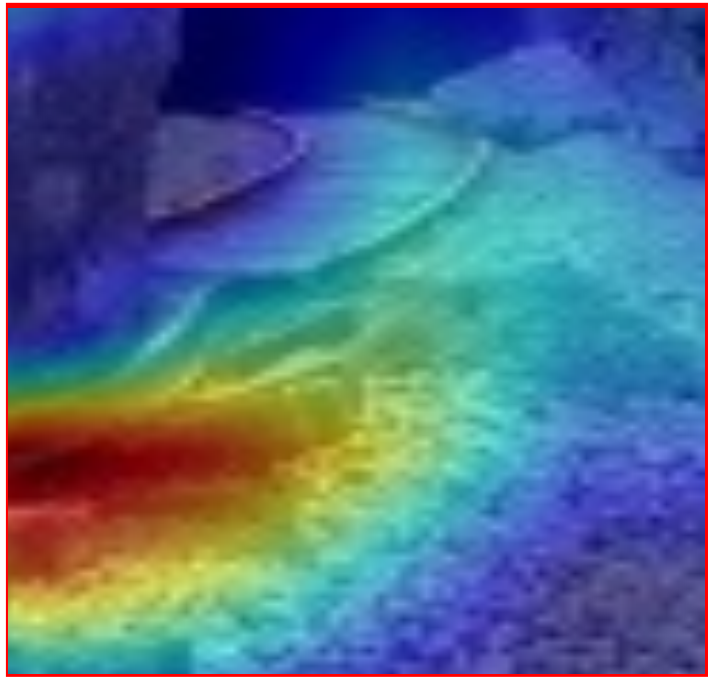


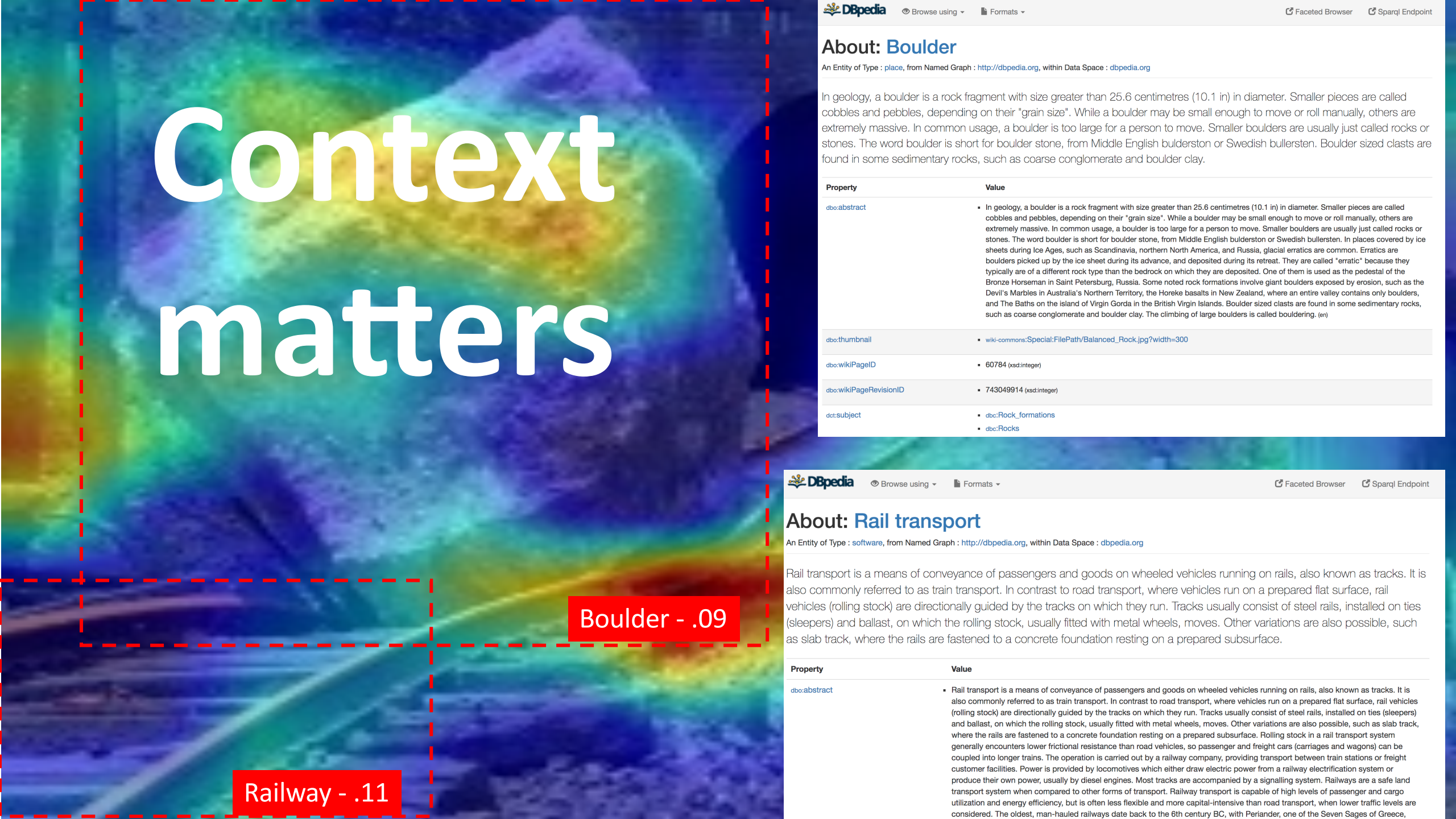
DBpedia | Browse using | Formats | Faceted Browser | Sparql Endpoint

dbo:wikiPageID	352327 (xsd:integer)
dbo:wikiPageRevisionID	734430894 (xsd:integer)
dct:subject	dbc:Sawmills dbc:Saws dbc:Ancient_Roman_technology dbc:Timber_preparation dbc:Timber_industry
http://purl.org/linguistics/gold/hypernym	dbr:Facility
rdf:type	owl:Thing dbo:ArchitecturalStructure
rdfs:comment	A sawmill or lumber mill is a facility where logs are cut into lumber. Prior to the invention of the sawmill, boards were rived (split) and planed, or more often sawn by two men with a whipsaw, one above and another in a saw pit below. The earliest known mechanical mill is the Hierapolis sawmill, a Roman water-powered stone mill at Hierapolis, Asia Minor dating back to the 3rd century AD. Other water-powered mills followed and by the 11th century they were widespread in Spain and North Africa, the Middle East and Central Asia, and in the next few centuries, spread across Europe. The circular motion of the wheel was converted to a reciprocating motion at the saw blade. Generally, only the saw was powered, and the logs had to be loaded and moved by hand. An early improvement was the developm (en)
rdfs:label	Sawmill (en)
owl:sameAs	wikidata:Sawmill dbpedia-cs:Sawmill dbpedia-de:Sawmill dbpedia-es:Sawmill

What is missing?

Lumbermill - .59





DBpedia Browse using Formats Faceted Browser Sparql Endpoint

About: Boulder

An Entity of Type : place, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

In geology, a boulder is a rock fragment with size greater than 25.6 centimetres (10.1 in) in diameter. Smaller pieces are called cobbles and pebbles, depending on their "grain size". While a boulder may be small enough to move or roll manually, others are extremely massive. In common usage, a boulder is too large for a person to move. Smaller boulders are usually just called rocks or stones. The word boulder is short for boulder stone, from Middle English bulderston or Swedish bullersten. Boulder sized clasts are found in some sedimentary rocks, such as coarse conglomerate and boulder clay.

Property	Value
dbo:abstract	▪ In geology, a boulder is a rock fragment with size greater than 25.6 centimetres (10.1 in) in diameter. Smaller pieces are called cobbles and pebbles, depending on their "grain size". While a boulder may be small enough to move or roll manually, others are extremely massive. In common usage, a boulder is too large for a person to move. Smaller boulders are usually just called rocks or stones. The word boulder is short for boulder stone, from Middle English bulderston or Swedish bullersten. In places covered by ice sheets during Ice Ages, such as Scandinavia, northern North America, and Russia, glacial erratics are common. Erratics are boulders picked up by the ice sheet during its advance, and deposited during its retreat. They are called "erratic" because they typically are of a different rock type than the bedrock on which they are deposited. One of them is used as the pedestal of the Bronze Horseman in Saint Petersburg, Russia. Some noted rock formations involve giant boulders exposed by erosion, such as the Devil's Marbles in Australia's Northern Territory, the Horeke basalts in New Zealand, where an entire valley contains only boulders, and The Baths on the island of Virgin Gorda in the British Virgin Islands. Boulder sized clasts are found in some sedimentary rocks, such as coarse conglomerate and boulder clay. The climbing of large boulders is called bouldering. (en)
dbo:thumbnail	▪ wiki-commons:Special:FilePath/Balanced_Rock.jpg?width=300
dbo:wikiPageID	▪ 60784 (xsd:integer)
dbo:wikiPageRevisionID	▪ 743049914 (xsd:integer)
dct:subject	▪ dbc:Rock_formations ▪ dbc:Rocks

DBpedia Browse using Formats Faceted Browser Sparql Endpoint

About: Rail transport

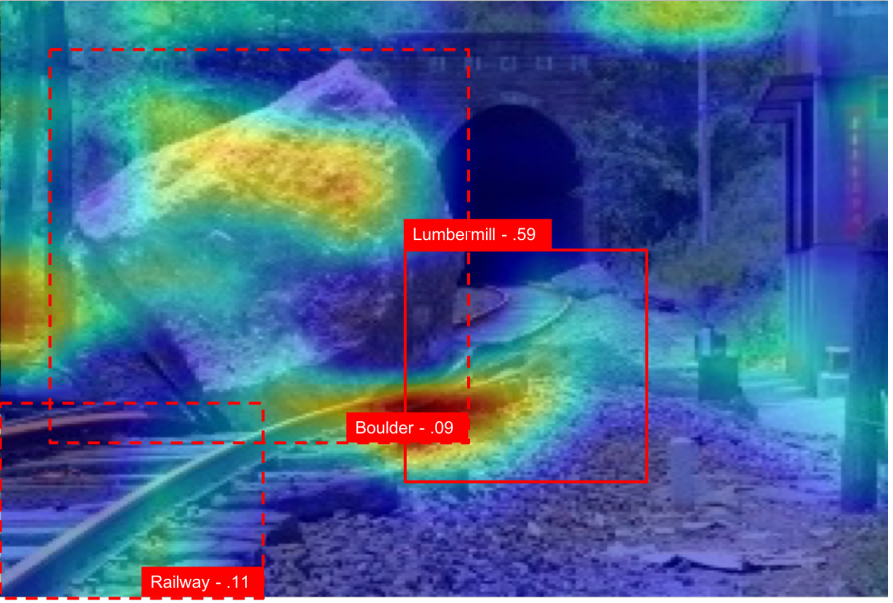
An Entity of Type : software, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

Rail transport is a means of conveyance of passengers and goods on wheeled vehicles running on rails, also known as tracks. It is also commonly referred to as train transport. In contrast to road transport, where vehicles run on a prepared flat surface, rail vehicles (rolling stock) are directionally guided by the tracks on which they run. Tracks usually consist of steel rails, installed on ties (sleepers) and ballast, on which the rolling stock, usually fitted with metal wheels, moves. Other variations are also possible, such as slab track, where the rails are fastened to a concrete foundation resting on a prepared subsurface.

Property	Value
dbo:abstract	▪ Rail transport is a means of conveyance of passengers and goods on wheeled vehicles running on rails, also known as tracks. It is also commonly referred to as train transport. In contrast to road transport, where vehicles run on a prepared flat surface, rail vehicles (rolling stock) are directionally guided by the tracks on which they run. Tracks usually consist of steel rails, installed on ties (sleepers) and ballast, on which the rolling stock, usually fitted with metal wheels, moves. Other variations are also possible, such as slab track, where the rails are fastened to a concrete foundation resting on a prepared subsurface. Rolling stock in a rail transport system generally encounters lower frictional resistance than road vehicles, so passenger and freight cars (carriages and wagons) can be coupled into longer trains. The operation is carried out by a railway company, providing transport between train stations or freight customer facilities. Power is provided by locomotives which either draw electric power from a railway electrification system or produce their own power, usually by diesel engines. Most tracks are accompanied by a signalling system. Railways are a safe land transport system when compared to other forms of transport. Railway transport is capable of high levels of passenger and cargo utilization and energy efficiency, but is often less flexible and more capital-intensive than road transport, when lower traffic levels are considered. The oldest, man-hauled railways date back to the 6th century BC, with Periander, one of the Seven Sages of Greece,

XAI Thales Platform

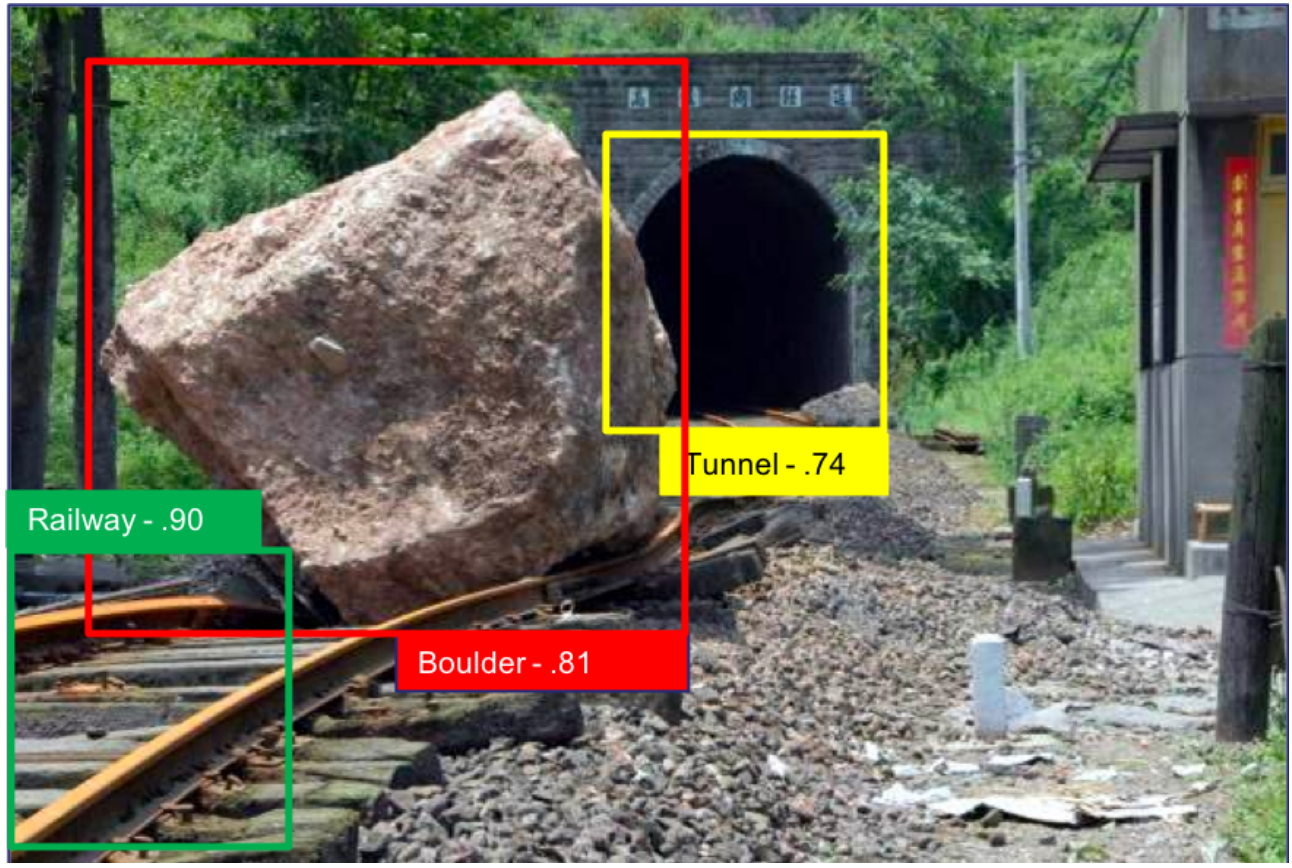
- Higher accuracy with no intensive fine-tuning
- Human interpretable explanation
- Running on the edge at inference time



- **Hardware**: **High performance, scalable, generic** (to different FGPA family) **& portable** CNN dedicated **programmable** processor implemented on an FPGA for **real-time embedded inference**

- **Software**: Knowledge graph extension of object detection

Transitioning

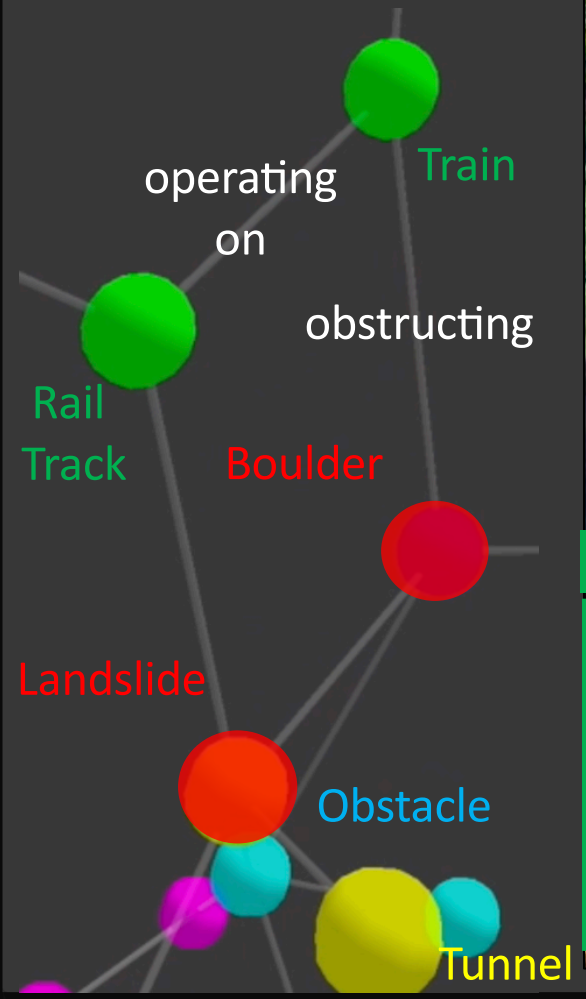


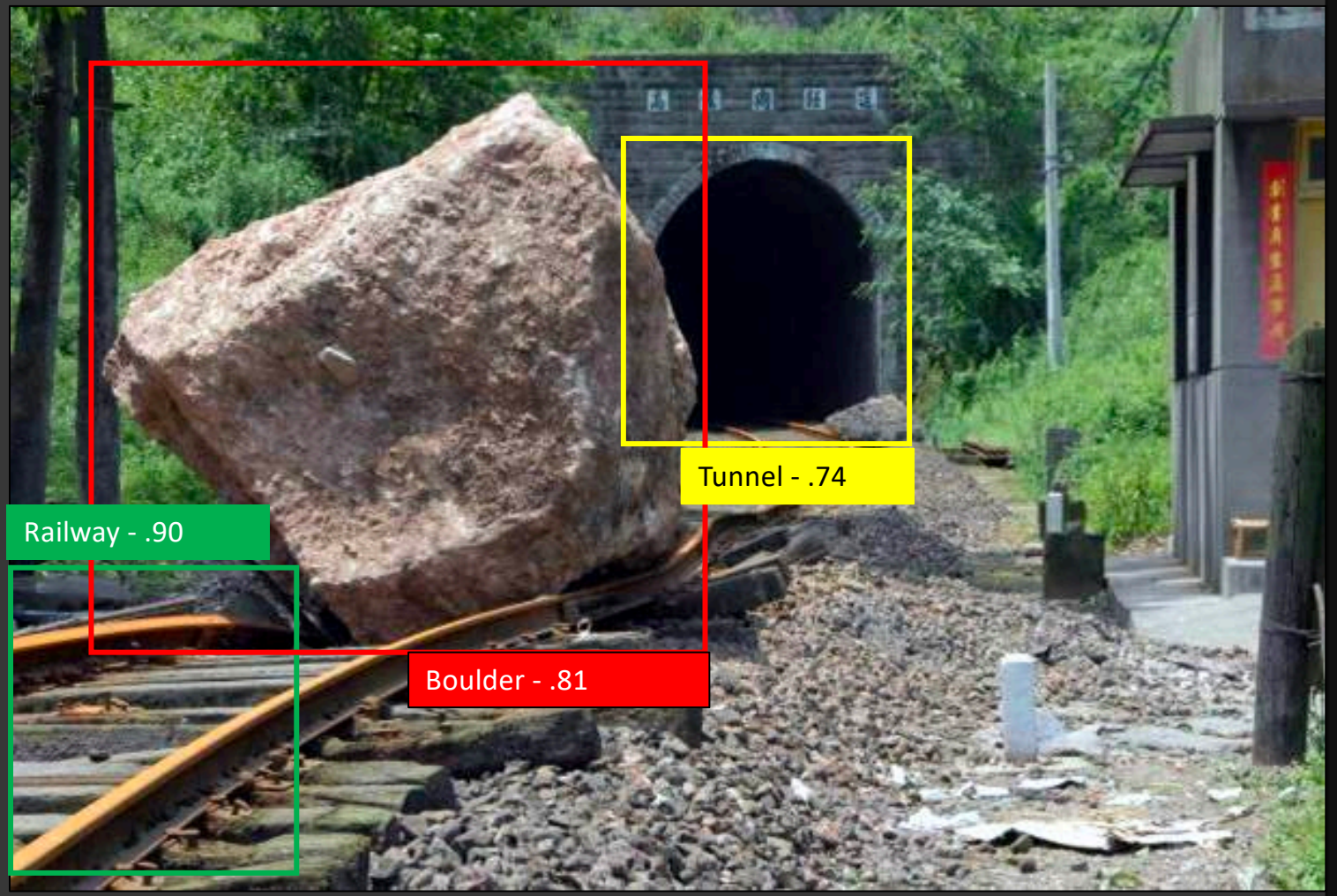
This is an **Obstacle: Boulder** obstructing the train: XG142-R on **Rail_Track** from City: Cannes to City: Marseille at **Location: Tunnel VIX** due to **Landslide**

EXPLANATIONS

Lime

ResNet50 image classifier





Conclusion

Conclusion

- Explainable AI is motivated by **real-world applications in AI**
- Not a new problem – a reformulation of past research challenges in AI
- Knowledge graphs should be foundational for XAI
- But they are facing challenges related to their integration (data mapping)
- **Many industrial applications already – crucial for AI adoption in critical systems**

Why do we Need Knowledge Graphs to Achieve XAI?

Because this is not an explanation from an intelligent system

This is even not interpretable, and then not actionable

Open Research Questions for the Semantic Web / Knowledge Graph Community

- [Data] Machine learning experts do not buy the **data – knowledge mapping**
- [Explanation] There is ***no agreement*** on ***what an explanation is***
- [Explanation] There is ***not a formalism*** for ***explanations (neither model nor output)***
- [Model] *There is very limited work in* ***machine learning modules composability –*** *and* ***none from a semantics perspective***
- [Model] ***There is no work on describing and representing models***
- [Model] What are **disentangled representations** and how can its factors be quantified and detected?
- [Human-in-the-loop] There is ***no work*** that seriously addresses the problem of ***quantifying*** the grade of ***comprehensibility*** of an explanation for humans