INSIGHTFUL EXPLANATIONS WITH KNOWLEDGE

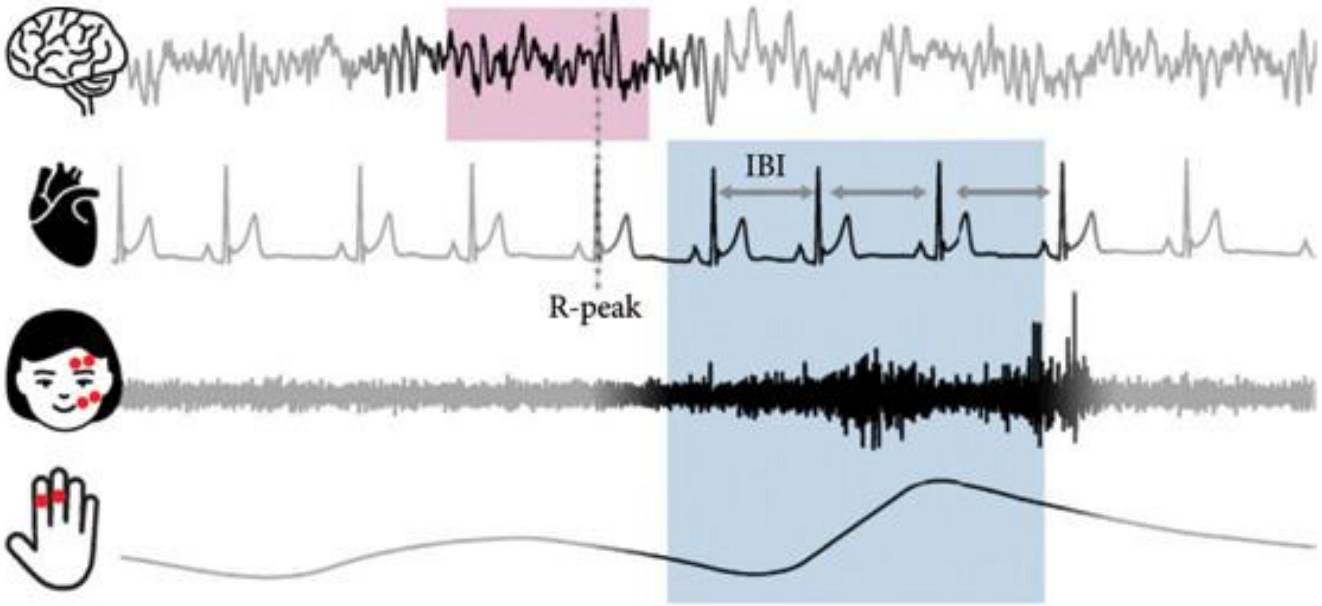
Krzysztof Kutt, PhD
Knowledge in AI Systems
WFAIS UJ

ML/DL WITHOUT KNOWLEDGE

Let's start with a tragedy

EMOTION PREDICTION FROM PHYSIOLOGICAL SIGNALS

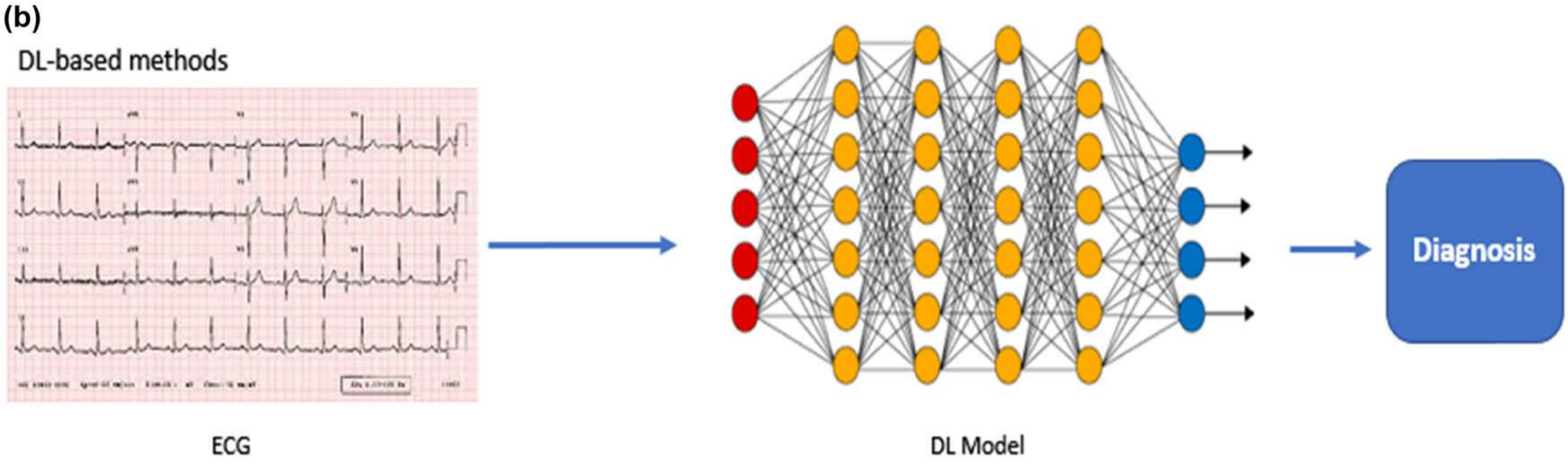
- *Input:* raw signals, e.g., electroencephalography (EEG), electrocardiography (ECG), electrodermal activity (EDA)
- *Output:* emotion-related label
- *Task:* create a model that predicts label based on raw signals



Source: H. A. Hamzah & K. K. Abdalla (2024), [EEG-Based Emotion Recognition Datasets for Virtual Environments: A Survey](#).

PURE DL-BASED APPROACH

- Simply put everything into some deep learning (e.g., with convolutional networks)
- The approach can be found even at top scientific AI/ML conferences (!)
- It may work, but…



Source: A. A. Rawi *et al.* (2023), Deep learning models for multilabel ECG abnormalities classification.

DOMAIN EXPERTS ARE THE EXPERTS

They have the knowledge

EMOTION PREDICTION FROM PHYSIOLOGICAL SIGNALS

Why do only ignorant people do this "pure DL way"?

- Physiological signals are by definition **noisy**
- **Raw** signal values are **meaningless** - what is more important are changes from baseline, differences over time, other characteristics related to changing body activity (heart, brain, etc.)

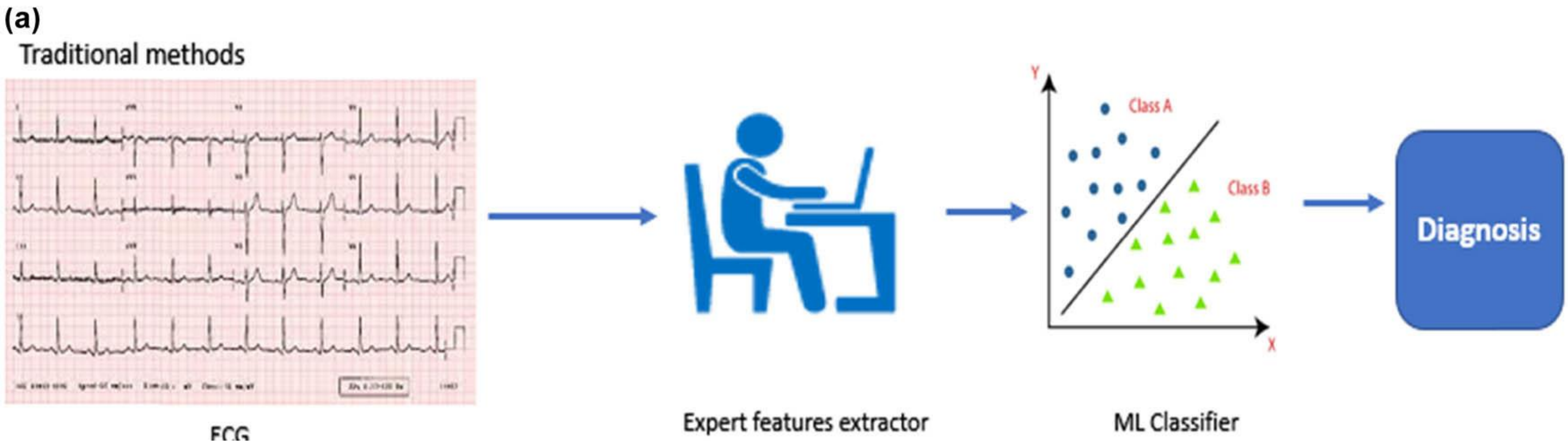
So, **yes, DL could extract this**, but it would take a huge amount of computing power and a huge amount of data to do so

But:

- Physiological signals have been analysed for many decades and we there are **textbooks** ready that tell you how to **filter** the signal, **which features are relevant**, what they mean, **...**
- We have ready-to-use libraries & tools that extract these features!

DOMAIN KNOWLEDGE-AWARE APPROACH

- Filtering and features extraction can be done automatically
- The inputs to ML/DL are meaningful features, not raw noise signals, so we are closer to the result



Source: A. A. Rawi *et al.* (2023), [Deep learning models for multilabel ECG abnormalities classification]().

NOT ONLY EMOTION PREDICTION

Mental health

- Stress
- Anxiety monitoring

Healthcare

- Sleep disorders
- Epilepsy

HCI

- BCI
- Adaptive UX

Security

- Lie detection
- Biometric auth

Workplace

- Fatigue monitoring
- Ergonomics

Marketing

- Neuromarketing
- Preference analysis

DL VS KNOWLEDGE-AWARE APPROACH

WE WANT TO UNDERSTAND

WE WANT TO UNDERSTAND – XAI METHODS

- SHAP
- LIME
- Permutation Importance
- Partial Dependence Plot
- Morris Sensitivity Analysis
- Accumulated Local Effects (ALE)
- Anchors
- Contrastive Explanation Method (CEM)
- Counterfactual Instances
- Integrated Gradients
- Global Interpretation via Recursive Partitioning (GIRP)
- Protodash
- Scalable Bayesian Rule Lists
- Tree Surrogates
- Explainable Boosting Machine (EBM)

Source: F. Dallanoce (2022), Explainable AI: A Comprehensive Review of the Main Methods.

WE WANT TO UNDERSTAND – XAI METHODS

Yes, we have a lot of XAI methods, but:

- we will **not** understand a model based **entirely on abstract deep features** (= we need some meaningful domain knowledge-based features)
- a situation in which only a few features are the most relevant and clearly explain the model is **wishful thinking** (= even with meaningful domain knowledge-based features, the explanation may be difficult to understand)

WE WANT TO UNDERSTAND – XAI METHODS

Yes, we have a lot of XAI methods, but:

- we will **not** understand a model based **entirely on abstract deep features** (= we need some meaningful domain knowledge-based features)
- a situation in which only a few features are the most relevant and clearly explain the model is **wishful thinking** (= even with meaningful domain knowledge-based features, the explanation may be difficult to understand)

[in affective computing and many other disciplines] there is domain knowledge about features, their interrelationships, interpretation of their combinations, etc

MOTIVATIONAL EXAMPLE 1

R. Caruana *et al.*, Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission, in: 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015, pp. 1721–1730.

- Reported model predicts that **asthmatic patients** have a **lower** risk of dying from pneuomina
- *But:* the doctor's medical expertise can reval that these patients were admitted directly to the Intensive Care Unit, receiving an aggressive care that indeed lowered their risk of death, but also caused incorrect machine-driven conclusions

Example discussed in: I. Tiddi & S. Schlobach (2022), Knowledge graphs as tools for explainable machine learning: A survey.

MOTIVATIONAL EXAMPLE 2

Clinicians <u>with</u> the intelligent agent explain the patient's case. Different types of explanations required at the different steps of the automated reasoning:

- "everyday explanations" for diagnosis
- "trace-based explanations" for planning the treatment
- "scientific explanations" to provide scientific evidence from existing studies
- "counterfactual explanations" to allow clinicians to add/edit information to view a change in the recommendation
- *[not in the paper, but also possible in such a case]* explanations for patient "with simple concepts"

Ontologies used to model knowledge to allow the AI system to automatically produce a wide range of explanations

Source: S. Chari *et al.* (2020), Explanation Ontology: A Model of Explanations for User-Centered AI.

XAI WITH KNOWLEDGE

Knowledge graph-based XAI

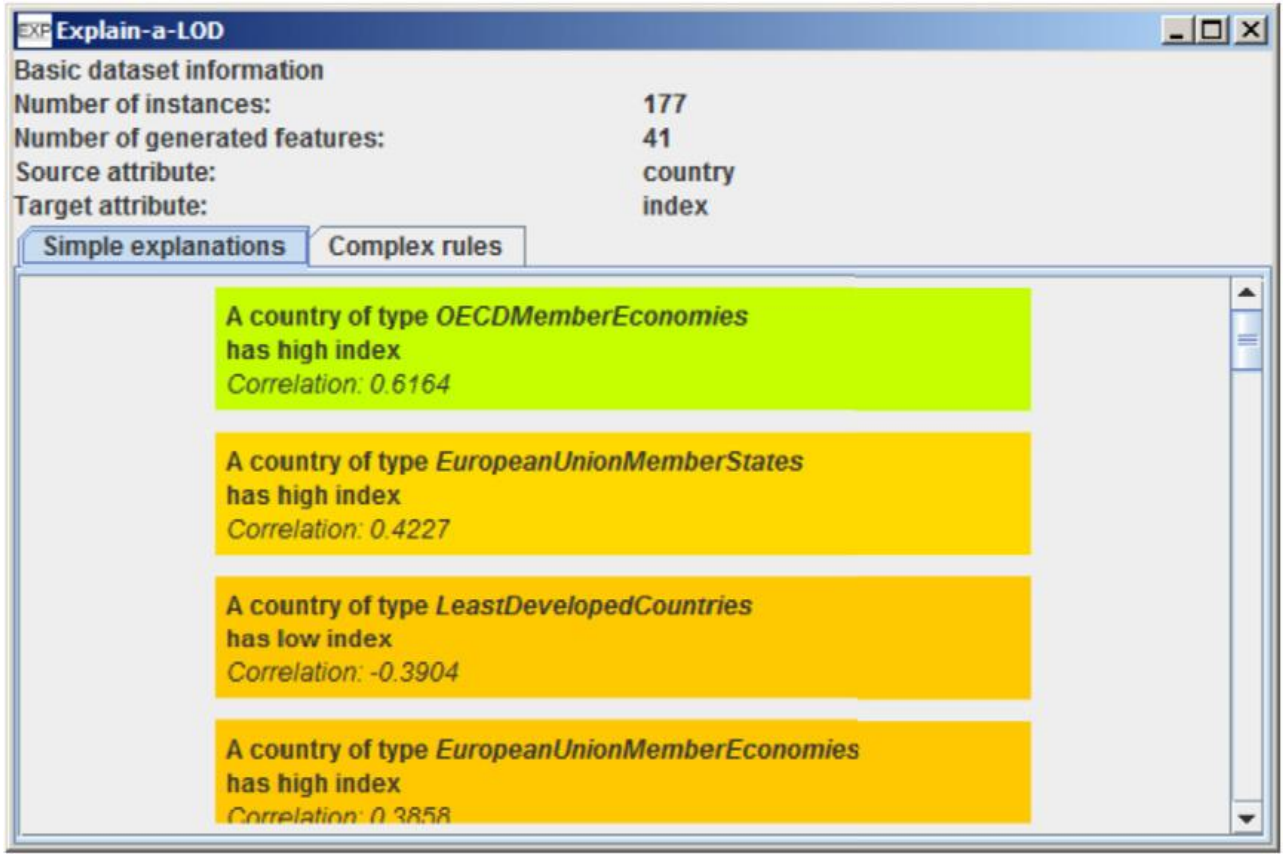
RULE-BASED ML

explain($\mathcal{Y}$): countries where males are more educated		
exp_i	F(%)	Time"
⟨*skos:exactMatch, dbp:hdiRank.≥.*"126"⟩	87.8	197"
⟨*skos:exactMatch, dc:subject.* **db:Category:Least_developed_countries**⟩	74.7	524"
⟨*skos:exactMatch, dbp:gdpPppPerCapitaRank.≥.*"89"⟩	68.3	269"
⟨*skos:exactMatch, dc:subject skos:broader.* **db:Category:Countries_in_Africa**⟩	67.1	540"
⟨*skos:exactMatch, dbp:populationEstimateRank.*"76"⟩	61.9	201"
⟨*skos:exactMatch, dbp:gdpPppRank.≥.*"10"⟩	59.1	235"

- Domain knowledge used to translate outputs of a neural network into symbolic knowledge
- In early approaches, properties and values from Linked Data were used directly to explain observations

Example discussed in: I. Tiddi & S. Schlobach (2022), [Knowledge graphs as tools for explainable machine learning: A survey]().

RULE-BASED ML

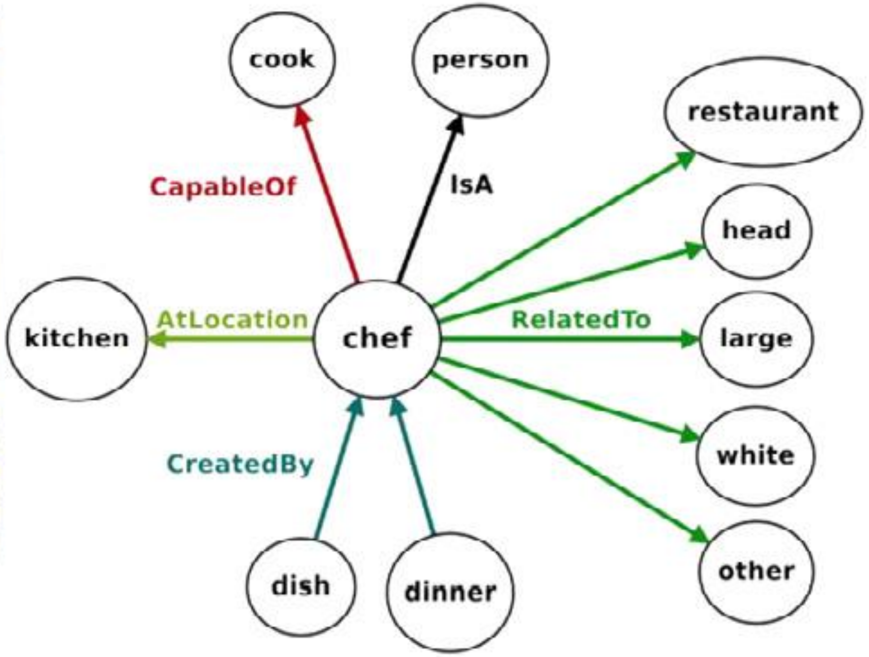


- Domain knowledge used to translate outputs of a neural network into symbolic knowledge
- In early approaches, properties and values from Linked Data were used directly to explain observations
- Used e.g. to explain statistical analyses to non-experts

Example discussed in: I. Tiddi & S. Schlobach (2022), [Knowledge graphs as tools for explainable machine learning: A survey](#).

IMAGE RECOGNITION

A **chef** getting ready to stir up some stir fry in the pan



- Domain knowledge is a user-friendly intermediate between the classifier and the end-user

Example discussed in: I. Tiddi & S. Schlobach (2022), [Knowledge graphs as tools for explainable machine learning: A survey]().

IMAGE RECOGNITION

- Domain knowledge is a user-friendly intermediate between the classifier and the end-user






$$\exists \text{contains.Window} \quad (1)$$
$$\exists \text{contains.Transitway} \quad (2)$$
$$\exists \text{contains.SelfConnectedObject} \quad (3)$$
$$\exists \text{contains.Roadway} \quad (4)$$
$$\exists \text{contains.Road} \quad (5)$$
$$\exists \text{contains.LandTransitway} \quad (6)$$
$$\exists \text{contains.LandArea} \quad (7)$$
$$\exists \text{contains.Building} \quad (8)$$
$$\forall \text{contains.}\neg\text{Floor} \quad (9)$$
$$\forall \text{contains.}\neg\text{Ceiling} \quad (10)$$

Example discussed in: I. Tiddi & S. Schlobach (2022), Knowledge graphs as tools for explainable machine learning: A survey.

RECOMMENDER SYSTEMS

Terminator 2: Judgment Day (1991)

★★★★★

Transformers: Revenge of the

We guess you would like to watch **Terminator 2: Judgment Day (1991)** more than **Transformers: Revenge of the Fallen (2009)** because you may prefer:

- (subject) 1990s science fiction films
- (subject) Science fiction adventure films
- (subject) Films using computer-generated imagery
- (subject) Drone films
- (subject) Cyberpunk films

over:

- (subject) Films set in Egypt
- (subject) Robot films
- (subject) Films shot in Arizona
- (subject) Ancient astronauts in fiction
- (subject) IMAX films

- Multi-edge paths extracted from the graph serve as explanations
- Example: explanations with *dc:subject*

Example discussed in: I. Tiddi & S. Schlobach (2022), Knowledge graphs as tools for explainable machine learning: A survey.

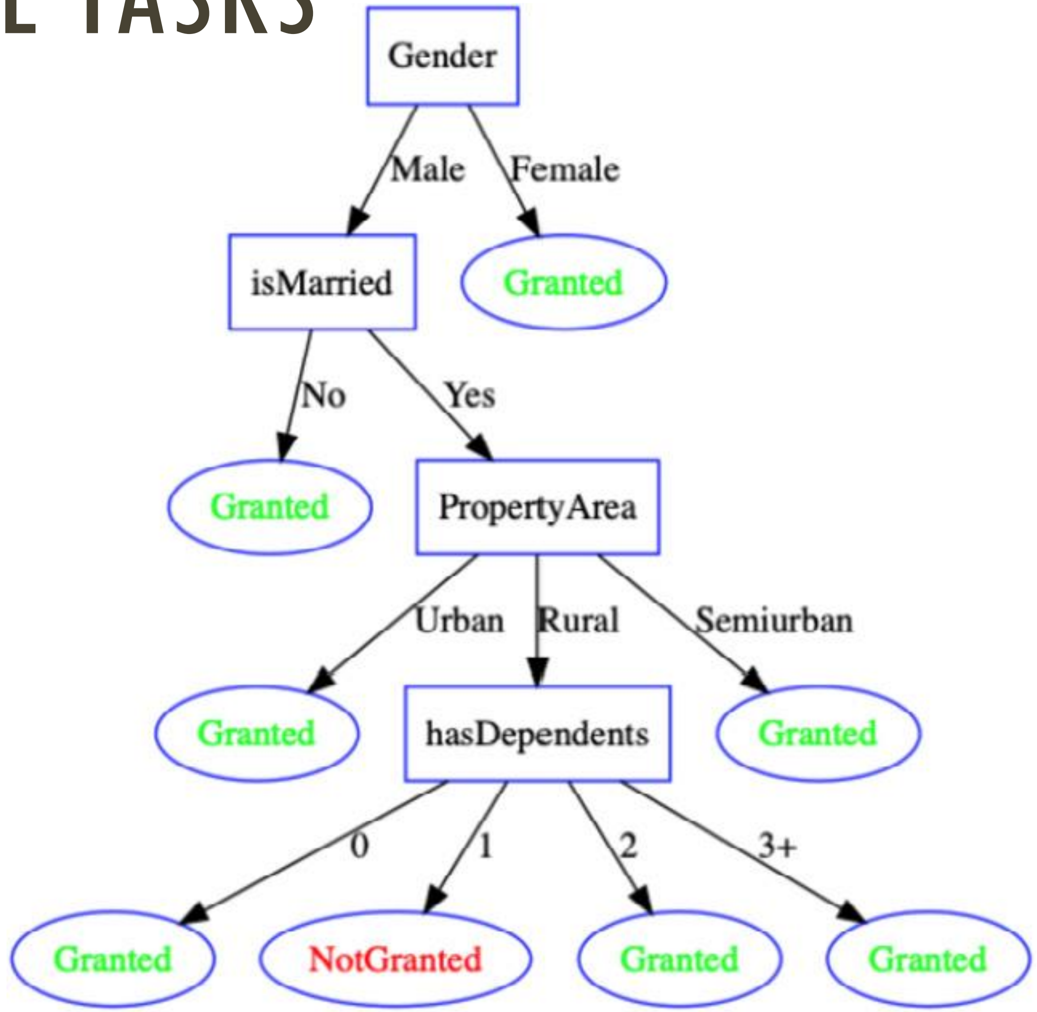
NATURAL LANGUAGE APPLICATIONS

Question: What can the red object on the ground be used for ?
Answer: Firefighting
Support Fact: Fire hydrant can be used for fighting fires.

- Combine Wikidata, WordNet, ConceptNet and others to provide background knowledge in a variety of context, such as images, text, speech

Example discussed in: I. Tiddi & S. Schlobach (2022), Knowledge graphs as tools for explainable machine learning: A survey.

PREDICTIVE TASKS



- Link raw input data points to nodes of the graphs and retrieve additional information through graph navigation

Example discussed in: I. Tiddi & S. Schlobach (2022), Knowledge graphs as tools for explainable machine learning: A survey.

KNOWLEDGE GRAPHS FOR XAI: DO THEY WORK?

KNOWLEDGE GRAPHS FOR XAI: DO THEY WORK?

- **More understandable systems** with human-readable explanations, but with trade-off between complex structure and succinctness (most often the explanations should be designed per specific task / group of tasks)
- **More accurate systems** with large-scale knowledge graphs, but there is a need for efficient knowledge extraction methods
- **Provide causal and analogical reasoning,** but at the cost of computational efficiency

SUMMARY

We are coming to the end

SUMMARY IN ONE SENTENCE

XAI, ML and DL **with knowledge** are **better** than pure ML/DL approaches without domain knowledge.

OPEN CHALLENGES

- **Knowledge graph maintenance.** Need for high-quality cross-domain knowledge graphs. Need for efficient approaches for knowledge graph evolution at scale.
- **Identity management.** Needed to efficiently use the available information.
- **Automated knowledge extraction from graphs.** Need for new heuristics to identify correct portion information.
- **Human role?** Human-in-the-loop explanations?
- **From knowledge to meaning.** Need for complex narratives (created with semantic models) able to capture meaning of certain experiences as humans do.

THANK YOU FOR YOUR ATTENTION!

GEIST Research Group: https://geist.re/

Krzysztof Kutt: https://krzysztof.kutt.pl/

KEEP
CALM
AND
ASK
QUESTIONS!

keep-calm.net